# Investigating the Effects of Teaching and Learning Tools in Chemistry Education

Myriam St-Onge Carle

Thesis submitted to the University of Ottawa
in partial fulfillment of the requirements for the degree of
Doctorate in philosophy in Chemistry

Department of Chemistry and Biomolecular Sciences
Ottawa-Carleton Chemistry Institute
Faculty of Science
University of Ottawa

# Table of contents

**Table of contents** ..... **ii**

**List of figures** ..... **v**

**List of tables** ..... **x**

**Abstract** ..... **xi**

**Acknowledgments** ..... **xiii**

**List of contributions** ..... **xv**

**List of abbreviations** ..... **xvii**

**Statement of contributions** ..... **xviii**

**Chapter 1. Introduction** ..... **1**

1.1 References ..... 5

**Chapter 2. Theoretical frameworks** ..... **8**

2.1 Modern information processing theory ..... 8
2.2 Mental models ..... 10
2.3 Constructive alignment framework ..... 12
2.4 Reasoning ..... 14
2.5 References ..... 16

**Chapter 3. Methods** ..... **20**

3.1 Overviews of methods in chemistry education research ..... 20
3.2 Positional statement: Potential biases in my work ..... 23
3.3 Validity in qualitative research ..... 26
3.4 Methods used in my doctoral research ..... 29
3.5 References ..... 30

**Chapter 4. Molecular dynamic visualization: what are participants mental models of the sub-microscopic level?** ..... **33**

4.1 Problem and context: Students' mental model of the sub-microscopic level may not be scientifically accurate ..... 33
4.2 Literature review: Mental models of the sub-microscopic level ..... 33
4.3 Literature Review: Molecular dynamic visualization as a tool for learning ..... 37
4.4 Molecular dynamic visualizations ..... 38
4.5 Learning tool: Odyssey by Wavefunction ..... 41
4.6 Goals and research questions ..... 41
4.7 Manuscript in preparation: Investigating participants mental models of the sub-microscopic level ..... 43
4.8 Introduction ..... 44
4.9 Theoretical frameworks ..... 44

4.10 Research questions....46
4.11 Method....46
4.12 Results and discussion....51
4.13 Conclusions....69
4.14 Implications for research....70
4.15 Implications for teaching....71
4.16 Limitations....71
4.17 References....72

**Chapter 5. OrgChem101.com: Mastering the arrows online module: what are the learning gains after using the module?.... 79**

5.1 Context: Student's struggle to use the electron pushing formalism to predict reaction....79
5.2 Literature review: symbolic language of organic chemistry – the electron pushing formalism....79
5.3 University of Ottawa Organic Chemistry Curriculum: Electron Pushing Formalism first and patterns of reactivity....81
5.4 Learning tool: OrgChem101.com – Mastering the Arrows....83
5.5 Goals and Research Questions....84
5.6 Published contribution....85
5.7 Introduction....86
5.8 Theoretical framework....89
5.9 Research questions....90
5.10 Methods....90
5.11 Results and Discussion....95
5.12 Conclusions....104
5.13 References....107

**Chapter 6. Learning outcomes for Delocalization: How are they Intended, Enacted, and Achieved? .... 112**

6.1 Delocalization: an important but confusing concept....112
6.2 Context: It is unclear what students should know about delocalization....113
6.3 Students' conceptualization and skills using delocalization....114
6.4 Learning tool: LOs to obtain constructive alignment....116
6.5 Goals and research questions....117
6.6 Published contribution: What are the delocalization learning outcomes, and how are they enacted....118
6.7 Introduction....120
6.8 Theoretical Framework....122
6.9 Goals and research questions....125
6.10 Methods....125
6.11 Results and discussion....128
6.12 Conclusions....143
6.13 Limitations....144
6.14 Implications for teaching and learning....144

6.15 Implications for research ..... 145
6.16 Acknowledgements ..... 145
6.17 Submitted contribution: How are the learning outcomes being achieved ..... 146
6.18 Introduction ..... 148
6.19 Research questions ..... 149
6.20 Theoretical frameworks ..... 150
6.21 Methods ..... 152
6.22 Results and discussion ..... 158
6.23 Conclusions ..... 185
6.24 Limitations ..... 187
6.25 Implications for research ..... 187
6.26 Implications for teaching ..... 188
6.27 References ..... 189

**Chapter 7. Conclusions ..... 196**

7.1 Summary of findings from each study ..... 196
7.2 Future directions ..... 201
7.3 Implications for educators ..... 207
7.4 Conclusion ..... 209
7.5 References ..... 210

**Appendix 1: "Molecular dynamic visualization" (Chapter 4) supplemental information ..... 216**

A1.1 Worksheet ..... 216
A1.2 Themes for randomness mental models ..... 217
A1.3 Types of learning and cognitive dissonance ..... 218

**Appendix 2: "OrgChem101.com" online module (Chapter 5) supplemental information ..... 219**

A2.1 Cohort I study ..... 219
A2.2 Coding scheme ..... 220

**Appendix 3: Synthesis of delocalization LO and Exam Analysis (Chapter 5) supplemental information ..... 226**

A3.1 Which term is appropriate: resonance or delocalization? ..... 226
A3.2 LOs and the percentage of questions related to each ILO ..... 227
A3.3 Example of Taught-Practiced-Assessed (TPA) coding ..... 228
A3.4 Exam analysis coding scheme for RQ1: LO Achievements ..... 230
A3.5 Expected answer for each question analyzed ..... 237
A3.6 Reliability for coding of the exams ..... 241
A3.7 Assessing LO2 when the starting material was incorrect ..... 242
A3.8 Type of answer for reasoning in question 3 ..... 243

# List of figures

Figure 1.1. Johnstone's triangle: the chemistry triplet (Johnstone, 2000) ..... 2
Figure 1.2. The three learning tools used in this thesis, with the associated level of the triplet ... 3
Figure 2.1. Representation of the modern information processing theory ..... 9
Figure 2.2. Schematic representation of how people build and use mental models ..... 12
Figure 2.3. Constructive alignment ..... 13
Figure 2.4. Format of a LO ..... 13
Figure 3.1. General flow of thematic analysis ..... 22
Figure 3.2. Example of validity and reliability. Each red point represents a data point obtained that was aimed at the bullseye ..... 25
Figure 3.3. Validity and reliability types along with the evidence required. ..... 26
Figure 4.1. Graphical relationships of the type of representation ..... 34
Figure 4.2. Deterministic and probabilistic mental models ..... 36
Figure 4.3. Schematic representation of mental models ..... 45
Figure 4.4. Overview of the study ..... 47
Figure 4.5. Screenshot of the simulation viewed by the students. Left: Acetone and water mixing, right: Fluorine and $NO_2$ reacting inside the cell. ..... 49
Figure 4.6. Frequency of static, process, and particles in motion mental models as a percentage of all mental models used (N = 167) ..... 52
Figure 4.7. Overview of the mental models of the sub-microscopic level. The size of the circles does not have meaning. ..... 53
Figure 4.8. Mental models of motion with associated codes and representative quotes ..... 54
Figure 4.9. Mental models of collisions and associated child codes and representative quotes. 56
Figure 4.10. Mental models that participants displayed related to collisions from the pre-interview to the post-interview. ..... 58
Figure 4.11. Mental models related to the amount of particles present compared to (a) static vs dynamic and (b) pre vs post. ..... 59
Figure 4.12. Participants at the junction of two mental models ..... 60
Figure 4.13. Example from P8 of a change in drawing, with the associated quote that they visualize the sub-microscopic level differently than drawn. ..... 61
Figure 4.14. Post-test drawing of P3 using a unique representation, with an associated quote. 62
Figure 4.15. Deterministic and probabilistic mental models ..... 63

Figure 4.16. Example and participant quotes of deterministic and teleological explanations (left, blue) and of probabilistic explanations (right, red). .....64

Figure 4.17. Mental models of participants about randomness .....65

Figure 4.18. P6's mental models before (left), using the simulation to learn (middle) and a more random mental model (right). .....69

Figure 5.1. An example of EPF using hydroxide and a carbonyl. .....80

Figure 5.2. LOs and key principles associated with the electron-pushing formalism.....87

Figure 5.3. Overview of the studies. .....91

Figure 5.4. Questions on the pre- and post-tests used in cohorts II and III, with Draw the Arrows questions (LO1) on the left with answers in red (bonds and electrons must be expanded if they are involved in the reaction step) and draw the Products questions (LO2) on the right with answers are in green. .....93

Figure 5.5. Post-test versus pre-test scores: Overall (left, blue circles), for LO1: Draw the arrows (middle, red triangles), for LO2: Draw the products (right, yellow squares). N = 103. Cohort. .....95

Figure 5.6. Score distributions per question on the pre- and post-tests (N = 103). Cohort II. .....96

Figure 5.7. Normalized learning gain scores, N = 103. Cohort II. .....97

Figure 5.8. Mapping instances found on the pre-test (orange) and post-test (blue) for each question. N = 103. Cohort II. .....98

Figure 5.9. Example of successful mapping used on the post-test, Question 8. .....99

Figure 5.10. Normalized learning gains for the intervention and control groups. Cohort II. .....100

Figure 5.11. Test scores, grouped according to whether mapping was observed in the pre-test and post-test. Cohort II. .....101

Figure 5.12. For LO1, Draw the Arrows questions: Overview of correct answers, errors in arrows, and attempts between the pre-test (left) and the post-test (right). Each data point represents a single arrow, N = 1522. .....102

Figure 5.13. Common errors found in answers for questions associated with LO2: Draw the products. N = 103. Cohort II. .....103

Figure 5.14. Example of errors seen in Draw the arrows questions, specifically Question 2. .....103

Figure 5.15. Example of a missing atom error in a Draw the products question (LO2), specifically Question 5. .....103

Figure 6.1. Resonance structures of an enone (left) and its resonance hybrid (right) .....113

Figure 6.2. Benzene can be represented as two resonance structure (left) or with a circle to represent aromaticity (right) .....121

Figure 6.3. ILOS can inform teaching strategies, learning activities, and assessment to achieve the desired LOs. .....123

Figure 6.4. Method to analyze the taught, practiced, and assessed aspects of delocalization concepts in organic chemistry. ....127

Figure 6.5. The ten essentials LOs for the first two semesters of organic chemistry. ....129

Figure 6.6. Occurrence of the ten essential ILOs in the textbooks (N = 7). ....132

Figure 6.7. The average location in the textbooks where the ILOs are taught. Error bars denote the standard deviation. ....133

Figure 6.8. Two different meanings and uses of curved arrows. ....134

Figure 6.9. Percentage of questions related to the LOs from all delocalization questions (N = 1548). "Other" represents any delocalization questions that do not fully address one of the 10 essential LOs. ....137

Figure 6.10. Distribution of delocalization question (%) in each chapter from the seven textbooks. N = 1548 ....139

Figure 6.11. The percentage of delocalization questions (N=1548) by subject. The top graph (a) is the traditional curriculum for Organic Chemistry I and the bottom graph (b) is the traditional curriculum for Organic Chemistry II. ....140

Figure 6.12. Percent of delocalization questions on assessments (N = 143) that are explicit (dark orange) and implicit (pale orange). The “Other” Category includes all questions that partially assess the LOs and any questions assessing “non-essential” ILOs. ....141

Figure 6.13.Percent of delocalization question (N = 143) related to each ILO on professors' assessments (N = 51) in Organic Chemistry I (dark orange) and Organic Chemistry II (pale orange) ....142

Figure 6.14. The ten essential LOs for delocalization. Reproduced with permission from Carle and Flynn, 2020 ....149

Figure 6.15. Twelve questions assessing ten LOs. Full answers can be found in the Supporting Information. ....154

Figure 6.16. LO achievement across all questions. Green = completely achieved, with justification, Yellow = partially achieved, had the correct answer but no justifications. Justifications were not required for the answers seen as partially achieved. ....159

Figure 6.17. Achievement of criteria for LO2 for each of the questions that assess LO2. Filled bars represent explicit questions, the dotted filled bar represents the implicit question, and lined-filled bars represent the mechanistic questions. ....161

Figure 6.18. The expected resonance structures for the five questions related to LO2 (Draw).163

Figure 6.19. Resonance structures for Question 12, missing two minor contributors. ....164

Figure 6.20. Comparison of errors in drawing the resonance structures. ....165

Figure 6.21. Modes of reasoning identified in answers to Question 1 (N = 280). ....168

Figure 6.22. Connection between achievement of LO2 and LO4 in Question 1. ....170

Figure 6.23. Hybridization labels for Question 1d (N = 284). The “other” category represents answers that are not represented by the three other levels. .....171

Figure 6.24. Answers given for Q5 (OCI, N = 288) and Q7 (OCII, N = 299). Cycles that are circled in green are aromatic, cycles that are squared in purple are anti-aromatic, and the remainders are non-aromatic. The percentage below or next to each cycle is the percentage of answers with that cycle correctly labelled. .....172

Figure 6.25. Distribution of students’ scores depending on whether they used a strategy on Question 5, N = 288 (top) and Question 7, N = 296 (bottom). .....173

Figure 6.26. Strategies used based on type of ring. .....174

Figure 6.27. Percentage of students who circled each site in Question 10. .....176

Figure 6.28. Frequency of correct answers for Question 10 (LO8) for answers with strategies and break down per strategy. .....177

Figure 6.29. Question 2 (acid–base): Evidence used to justify the claim (N = 280). .....177

Figure 6.30. Percentage of answers with links between concepts (n = 155). The percentage in parentheses is out of all answers, including those who did not identify resonance (N = 286). .....178

Figure 6.31. Modes of reasoning for Question 2 (acid/base). (N = 286). .....180

Figure 6.32. Common answers for Question 10. The correct answer is in green. .....181

Figure 6.33. Frequency of students mentioned a concept when not explicitly asked. (a) Question 3, OCI, N = 288, (b) Question 8, OCII, N = 296. Orange: used the concept and achieved the LO, blue: mentioned the concept but did not achieve the LO. .....182

Figure A2.1. Coding scheme for Draw the arrow questions. Each arrow was awarded 1 point. 221

Figure A2.2. Coding scheme for question 5: Draw the products. .....222

Figure A2.3. Coding scheme for question 7: Draw the products. .....222

Figure A2.4. Coding scheme for question 6: Draw the products. .....223

Figure A2.5. Coding scheme for question 8: Draw the products .....223

Figure A3.1. Expected answer for the Question 1. .....237

Figure A3.2. Expected answer for Question 2. .....238

Figure A3.3. Expected answer for Question 3. .....238

Figure A3.4. Expected answer for Question 4. The correct answer is in blue. .....239

Figure A3.5. Expected answer for Question 4. The most stable ion is highlighted in blue. .....239

Figure A3.6. Expected answer for Question 5. Blue = aromatic, Red = anti-aromatic, black = non-aromatic. .....239

Figure A3.7. Expected answer for Question 7. Blue = aromatic, Red = anti-aromatic, black = non-aromatic. .....239

Figure A3.8. Expected answer for Question 8. .....240

Figure A3.9. Expected answer for Question 9. The two structures are the expected structures to be drawn. A (blue) is the major product formed. .....240

Figure A3.10. Expected answer for question 10. The most basic atom is in blue. .....240

Figure A3.11. Expected answer for question 11. The most nucleophilic atom is in blue. .....240

Figure A3.12. Expected answer for question 12. .....241

Figure A3.13. Question 2: Common error (18%) of drawing the wrong charge on the resonance structures, despite drawing the correct curved arrows. .....243

Figure A3.14. Levels of reasoning in Question 1 (resonance), N = 280. .....243

# List of tables

Table 2.1. Description of the level of reasoning from Sevian and Talanquer (2014). .....15

Table 3.1. Differenced between qualitative and quantitative research.....21

Table 3.2. Types of validity in qualitative data as outlined by Maxwell (2012).....27

Table 3.3. Specific approach to obtain evidence for validity in qualitative research, aligned with Maxwell (2012) types of validity.....27

Table 5.1. t-test statistics from comparing the pre-test and post-test scores for the overall score and scores on questions related to the individual LOs.....96

Table 5.2. Mann-Whitney test for comparison of pre-test scores and post-test scores for the intervention and the control groups. .....98

Table 5.3. Mann-Witney tests to compare the normalized learning gains from the control and the intervention groups. .....98

Table 6.1. Selected questions aligned with the LOs. .....153

Table 6.2. General criteria for LO achievement.....155

Table 6.3. Criteria used to identify modes of reasoning in Question 1 and Question 2.....157

Table 6.4. Examples of modes of reasoning for Question 3. .....179

Table A1.1. Themes for randomness.....217

Table A1.2. Types of cognitive dissonance .....218

Table A2.1. Coding scheme for errors in Draw the arrows questions. .....224

Table A2.2. Coding scheme for the errors of LO2 Draw the products.....225

Table A3.1. Percentage of all delocalization questions (N = 1548) related to the 33 ILOs. .....228

Table A3.2. Coding examples of how questions were coded. LOs are listed in Table 1. .....229

Table A3.3. General criteria to achieve the LOs .....230

Table A3.4. Coding scheme for LO2.....231

Table A3.5. Coding scheme of student reasoning for LO3 (Assess).....232

Table A3.6. Coding scheme for LO4.....233

Table A3.7. List of concepts coded for LO8 (Acid/base) - the acid–base question.....234

Table A3.8. Mode of reasoning for LO8 (Acid–base) – question 2 .....235

Table A3.9. Coding scheme for LO10 (Reaction) question - EAS questions.....236

Table A3.10. Extra information coded for EAS type questions.....236

Table A3.11. Coding for the strategies used.....237

Table A3.12. Reliability of coding. Number in parentheses are the unsuccessful first rounds. .242

# Abstract

My research focused on three projects: (1) investigation of the mental models of students of the microscopic world using a molecular dynamic visualization, (2) evaluation of an online module on students' skills related to electron-pushing formalism, and (3) a two-part investigation of how 10 essential learning outcomes (LOs) about delocalization were intended, enacted, and achieved.

**Project 1: Exploring participants mental models of the sub-microscopic level after viewing a molecular dynamic visualization.** The effect of two molecular dynamic simulations on students' mental models about motion, collisions, and probabilistic thinking was investigated via a qualitative study. We administered a worksheet and interviewed the participants both before and after they viewed the visualizations. The analysis showed that (1) participants all had a motion mental model, (2) participants used different mental models depending on the situation, (3) participants had conflicting mental models of randomness of the sub-microscopic level, and (4) participants experienced cognitive dissonance when viewing the simulation.

**Project 2: Evaluation of OrgChem101.com online module of students' skills using EPF.** We investigated students' skills on the electron-pushing formalism after using an online learning module called "Organic Mechanisms: Mastering the Arrows" using a quantitative experimental method. There were significant learning gains between the pre- and post-test, especially with questions that asked students to draw the products of a reaction. After using the learning tool, students used more analysis strategies, such as mapping, attempted more questions, and made fewer errors.

**Project 3.1: Determining essential LOs for delocalization and how they are taught, practiced, and assessed.** The 10 LOs about delocalization (i.e., resonance) were determined from a textbook analysis then investigated for how they are being *enacted*, meaning how they were taught, practiced, assessed. We have found that five themes emerged from the analysis: (1) Several of the essential intended LOs we identified are not represented in the textbooks' teaching explanations, practice questions, or professors' assessments; (2) The concepts related to delocalization are often taught, practiced, and assessed without associated justifications; (3)

There is a large gap between when delocalization is taught and when it is used in context; (4) The link between delocalization and other concepts (e.g., reactivity) is not explicitly explained in most teaching materials; and (5) The language used around delocalization may be misleading (e.g. resonance, stability).

**Project 3.2: Investigating how the 10 essential delocalization LOs are achieved on summative examination.** We then analyzed how the students *achieved* the 10 essential LOs about delocalization on a summative assessment by analyzing 12 questions related to the concept. We found that students sometimes struggled to identify when delocalization could occur, that some of the LOs built on one another, and that some strategies (visualizing electrons, listing properties, and expanding the structures) more often led to the correct answer. We also found that when explicitly asked students in organic chemistry one was more successfully than in organic chemistry II and that the opposite occurs when asked within a mechanism. Our analysis of student reasoning showed that the dominant modes of reasoning were aligned with the related expectations and explanations in the course. When asked to justify the contribution of resonance structures to the resonance hybrid, most answers analyzed showed a descriptive mode of reasoning; when asked to explain why a given proton was more acidic than another, most answers contained relational and linear causal reasoning.

# Acknowledgments

This work was funded by the Social Science and Humanities Research Council (SSHRC) through a Joseph-Armand Bombardier (CGS-D) scholarship.

Thank you, Alison, for being a great mentor and friend. I have learnt so much from you, both professionally and personally. You have pushed me to become the best researcher (and person) I could be! I am forever grateful that I have had the chance to work with you and call you a friend.

I would like to also thank my Thesis Advisory Committee, Dr. Sean Barry, Dr. Chris Boddy, and Dr. Christine Tippett for your constant support and guidance. You have helped shape my research and helped me become a good scholar.

The best part of this this journey was the people I met and worked with. I want to thank every member of the Flynn Research Group for your help and guidance. There are many memories that I will forever cherish. I want to specifically thank Keith (for all those afternoons talking about our ideas), Amanda (you taught me so much about CER!) Jacky (the only person I always seem to run into while out for a walk), Alisha (for all those talks we had about everything!), Denzel (for all those talks about food), and Nick (for those times we spent climbing together). Also, thank you to Natalie, Nik Bodé, Zena, Nick Pilote, Roméo, Malek, Elly, Allison, Emily, Ahmed, Samira – you make the Flynn Research Group a great community!

I also want to thank my close friends who have supported me. Thank you, Michaela and Darcy, for making the pandemic better by playing all those games online. Meli, for all those fancy afternoon teas and for our Downton Abbey talks. Thank you, Meredith, for being the friend I can always count on to go for “just one beer”.

I am thankful to also have found a community and friends outside of my research. The varsity squash team really allowed me to relax and be active. Thank you to all my squash friends – Vinit, Taylor, Kabir, Eric, Julia, Caro, Miranda, Daniel. Thank you for keeping active and having fun!

My family has always been my rock. I cannot begin to thank you all enough. Mom and Dad - thank you for all of your support. Without the support you both provided me I wouldn't have achieved everything that I have so done so far. Also, my brother, his wife, and my three adorable nieces, you made me forget all the rough time during my PhD. I always love seeing pictures of the little ones!

Finally, Neil. You have been with me all the way, through the highs and lows. I love you. Thank you for being the best partner I could've had! I cannot wait to see what our life together brings, and I know it'll be great because we'll share it with each other!

All of you, and many more, have supported me in my journey and your support has allowed me to be here today. This thesis would not have been possible without all of you. Also, coffee… Thank you Starbucks for fueling my PhD.

Also thank you to Meeple and Bear, for being adorable cats who tolerates me.

# List of contributions

The following section outlines the contribution that arose from my PhD research. Listed are the peer-reviewed publications, along with conference presentation related to my PhD thesis.

**Publications**

- **Carle, M.S.,*** Deng, J., Flynn, A. B., (2022), Supporting students' reasoning skill development in chemistry arguments through constructive alignment, Invited contribution to "Reasoning in chemistry education". *Accepted.*
  *Co-first author with Deng, J.
- **Carle, M.S.,** Flynn, A.B. (2021), Investigation of participants multiple mental models of the sub-microscopic level and how external representation can lead to cognitive dissonance. *In prep.*
- **Carle, M.S.,** Issa, R., Pilote, N., Flynn, A.B., (2021) Ten essential delocalization learning outcomes: How well are they achieved? *ChemRxiv.* DOI:10.33774/chemrxiv-2021-c12b9-v2
- **Carle, M. S.,** Deng, J. M., Huang, D, Lapierre, K. R., Mesnic, N., Bodé, N., Featherstone, R., Ingram, Q., O'Connor, E. K., Roberge, S., Roy, K., Veilleux-Deschênes, J., Znotinas, A.; Flynn, A. B., (2019) Chapter 1: Engaging students as co-creators of course resources and learning experiences, *New direction of teaching and learning*, 2021, 63-76.
- **Carle, M.S.,** Flynn, A.B. (2020) Essential learning outcomes for delocalization (resonance) concepts: How are they taught, practiced, and assessed in Organic Chemistry? *Chemistry Education Research and Practice,* 2020, **21**, 622-637.
- **Carle, M. S.,** Visser, R., Flynn, A.B., (2020) Evaluating students' learning gains, strategies, and errors using OrgChem101's Module: Organic Mechanisms — Mastering the arrows, *Chemistry Education Research and Practice,* 2020, **21**, 582-596.

**Presentations**

- (Oral) Essential learning outcomes for delocalization (resonance) concepts: In what ways are they taught, practiced, assessed, and achieved in organic chemistry courses? American Society of Chemistry Spring Meeting 2022 (**ASC2022**), Online, March 20-25, 2022.
- (Oral) Investigating students' mental models before and after viewing a 2D simulation, Canadian Chemistry Conference and Exhibition 2021 (**CCCE2021**), Online, August 17, 2021.
- (Poster) How are 10 delocalization learning outcomes being enacted and achieved? Royal Society of Chemistry Twitter Conference 2021 (**RSCPoster 2021**), Online, March 3, 2021
- (Poster) Investigating how delocalization is taught, practiced, and assessed in organic chemistry, Method in Chemistry Education Research 2020 (**MICER 2020**), Online, June 11, 2020

- (Poster) Essential learning outcome for the concept of delocalization: how are they taught, practiced, and assessed? Royal Society of Chemistry Twitter Conference 2020 (**RSCPoster 2020**), Online, March 3, 2020
- (Oral) What do students need to know about resonance/delocalization? Ottawa-Carleton Chemistry Institute 2020 (**OCCI 2020**), Online, June 4, 2020.

  *Best 5 min presentation award
- (Oral) Designing a virtual reality (VR) experience for students to visualize a reaction, Canadian Chemistry Conference and Exhibition 2020 (**CCCE2020)**, Winnipeg, Manitoba, Canada, May 2020.

  *The conference was cancelled due to COVID-19
- (Poster) Systematic investigation of how delocalization is taught, practiced, and assessed in Organic Chemistry, Gordon Research Conference – Chemistry Education Research and Practice 2019 (**GRC-CERP 2019**), Lewiston, Maine, USA. June 16-21, 2019.
- (Poster) Systematic investigation of how delocalization is taught, practiced, and assessed in Organic Chemistry, Canadian Chemistry Conference and Exhibition 2019 (**CCCE 2019**), Quebec City, Quebec, Canada. June 3-7, 2019.
- (Oral) Studying the effectiveness of OrgChem101.com Mechanism Module on students' learning gains, Canadian Chemistry Conference and Exhibition (**CCCE 2019**), Quebec City, Quebec, Canada. June 3-7, 2019.

  *Best oral presentation award
- (Poster) OrgChem101.com Organic Mechanisms: Is it a useful tool to teach students the electron-pushing formalism? Royal Society of Chemistry Twitter Conference 2019 (**RSCPoster 2019**), Online, March 5, 2019.
- (Oral) Designing a Virtual Reality (VR) environment for students to visualize a reaction, Ottawa-Carleton Chemistry Institute 2019 (**OCCI 2019**), University of Ottawa, Ontario. Abstract #107, May 23, 2019.
- (Oral) OrgChem101.com: An OER to help learn Organic Chemistry, Open Education Recourses Exposition 2019 (**OER-Expo 2019**), University of Ottawa, Ottawa, Ontario, Canada. March 5, 2019.
- (Oral) Evaluation of how delocalization is taught, practiced, and assessed, Cross-Faculty Symposium of Science Education 2018, University of Ottawa, Ottawa, Ontario, Canada. November 3, 2018
- (Oral) What are the students' learning gains and experiences using Orgchem101.com mechanism module? Biennial Conference of Chemistry Education 2018 (**BCCE 2018**), Notre Dame University, South Bend, Indiana, USA. Presentation 345. July 29-August 2, 2018
- (Poster) OrgChem101.com Organic Mechanisms: Is it a useful tool to teach students the electron-pushing formalism? Canadian Society of Chemistry Conference 2018 (**CSC 2018**), University of Alberta, Edmonton, Alberta, Canada. Poster 181420. July 29-August 2, 2018.

  *Best poster award
- (Oral) What are the students' learning gains and experiences using Orgchem101.com mechanism module? Ottawa-Carleton Chemistry Institute (**OCCI 2018**), Carleton University, Ottawa, Ontario. May 22, 2018.

# List of abbreviations

| | |
|---|---|
| ASC | American Society of Chemistry |
| CCCE | Canadian Chemistry Conference and Exhibition |
| CER | Chemistry Education Research |
| CERP | Chemistry Education Research and Practice |
| COPUS | Classroom Observation Protocol for Undergraduate in STEM |
| DBER | Discipline Based Education Research |
| EAS | Electrophilic Aromatic Substitution |
| EPF | Electron Pushing Formalism |
| ILO | Intended Learning Outcomes |
| IPT | Information Processing Theory |
| IUPAC | International Union of Pure and Applied Chemistry |
| LO | Learning Outcome |
| LTM | Long Term Memory |
| MS | Mass Spectrometry |
| NMR | Nuclear Magnetic Resonance |
| OCCI | Ottawa-Carleton Chemistry Institute |
| OCI | Organic Chemistry I |
| OCII | Organic Chemistry II |
| OER | Open Education Resource |
| RQ | Research Question |
| RSC | Royal Society of Chemistry |
| TA | Teaching Assistant |
| WM | Working Memory |

# Statement of contributions

Chapter 4 is a manuscript written by Myriam S. Carle, Peter G. Mahaffy, and Alison B. Flynn. All authors were involved in the brainstormed the idea and designed the study. ABF and PGM provided guidance and feedback throughout the entirety of the project.

Chapter 5 is a manuscript written by Myriam S. Carle, Rebecca Visser, and Alison B. Flynn. RV and ABF designed the workshops and did the data collection and analysis for the pilot studies. The worksheet and protocol were designed by RV and ABF. ABF provided guidance on all phases.

Chapter 6 consists of two manuscripts. The first is written by Myriam S. Carle and Alison B. Flynn. ABF provided guidance and feedback throughout the entirety of the project.

The second manuscript in Chapter 6 was written by Myriam S. Carle, Roméo El Issa Jr, Nicholas Pilote, and Alison B. Flynn. REI performed the first round of coding on Question 3 and provided a preliminary code book for LO10 (Reaction) and LO2 (Draw – within a mechanism). NP performed the first round of coding on Question 2 and provided a preliminary code book for LO2 (Draw), LO7 – Stability, and LO8 (Acid—base). ABF provided guidance and feedback on all stages of the project.

# Chapter 1. Introduction

*"Learning is an enduring change in behaviour, or in the capacity to behave in a given fashion which results from practice or other form of experience" (Schunk, 2012, p. 3)*

Learning represents a change that remains and is built from experiences. Learning is interconnected with teaching, which is the creation of a learning environment that assists students in their learning (Floden *et al.*, 2001). Education research is the systemic gathering of evidence on student learning and teaching method with the goal to improve the learning experience (Creswell, 2012; American Educational Research Association, 2021). Discipline-Based Education Research is the field of education research in a specific domain, such as Chemistry Education Research (Henderson *et al.*, 2017). Chemistry Education Research aims to provide evidence-based practices towards improving teaching and learning of chemistry (National Research Council, 2015).

Chemistry can be difficult to learn because it occurs at a scale that is invisible, but chemistry is important to learn because it is related to many other disciplines and global issues (Balaban and Klein, 2006; Pazicni and Flynn, 2019). Understanding how particles we cannot see affect our daily lives can make chemistry difficult to learn. The chemistry triplet, or Johnstone's triangle addresses the fact that chemistry occurs on three different levels: the representational level, the sub-microscopic level, and the macroscopic level (Figure 1.1) (Johnstone, 1991; Taber, 2009; Talanquer, 2011). The representational level represents the language that chemists use to talk about chemistry. The macroscopic level represents what is visible to the naked eye and can be interacted with. The sub-microscopic level represents the scale that we cannot see nor interact with, in which atoms and molecules exist. Experts have the skills to readily understand and alternate between the three levels (Gilbert, 2008; Johnstone, 2010). However, students may struggle with alternating between the domains and may not be able to translate between domains (Kozma and Russell, 1997; Johnstone, 2000). To master chemistry, a person should be

able to understand and use all three levels and understand the relationship between the levels (Taber, 2009).

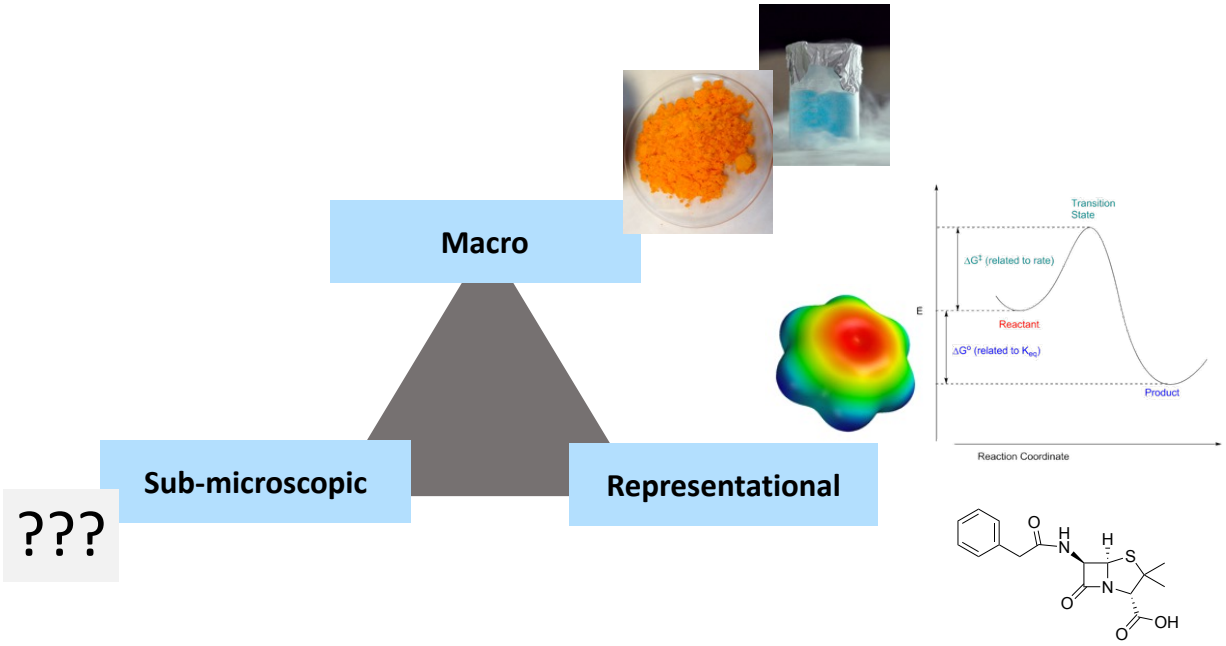



*Figure 1.1. Johnstone's triangle: the chemistry triplet (Johnstone, 2000)*

Many tools exist to help learners gain skills and knowledge about chemistry including visualizations, online modules, and textbooks. However, a tool may or may not have the intended effects and may lead to learners gaining alternate conceptions (Kelly and Jones, 2007; Tasker and Dalton, 2008), using the tools without understanding (Kelly and Jones, 2007; Kelly, 2014) or focusing on unintended aspects. Therefore, more evidence is required to explore the effects of these teaching tools.

My thesis focuses on three different tools and how they influence learners' skills of the symbolic level or visualization (i.e., mental models) of the sub-microscopic level of chemistry (Figure 1.2). The main research question for this thesis is: *How might learning tools (i.e., molecular dynamic visualization, learning outcomes, online module) help participants build mental models of chemistry?* The goal is to provide evidence on the effect of learning and teaching tools.

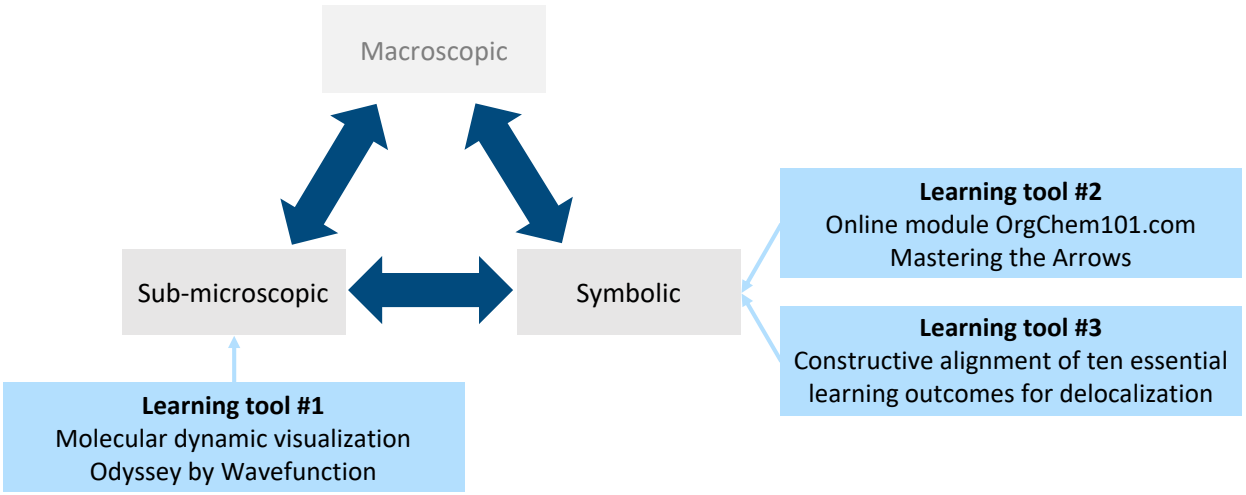



*Figure 1.2. The three learning tools used in this thesis, with the associated level of the triplet*

In the first project (Chapter 4), I explored participants' mental models of the sub-microscopic level after they viewed a molecular dynamic visualization. While chemical processes are dynamic and random, some students tend to have a static and non-random view of the sub-microscopic level (Adbo and Taber, 2009; Bongers *et al.*, 2019). Molecular dynamic visualizations are tools that can be used to encourage learners to develop a dynamic mental model of the molecular world (Höffler and Leutner, 2007). This project answered the following research question: *What are learners' mental models of the sub-microscopic level before and after viewing a molecular dynamic visualization?*

In my second project (Chapter 5), I evaluated an online learning module (OrgChem101.com Mastering the arrows) by looking at students' skills using the symbolic level. OrgChem101.com is a website that aims to provide students with the tools to succeed in three areas: nomenclature (Flynn *et al.*, 2014), acid-base chemistry (Stoyanovich *et al.*, 2015) and electron-pushing formalism (i.e., the curved arrows demonstrating electron movement). While the nomenclature module has been evaluated (Bodé *et al.*, 2016) and shown to have benefits to students, the other modules have not been assessed, therefore I evaluated the mechanistic module for students learning gains. The mechanistic module focuses on the symbolic aspect of the triplet, by teaching the electron pushing formalism—the language of organic chemistry. This project answered the following research question: *How might an online module affect students' skills of the symbolic level of organic chemistry?*

In the third project (Chapter 6), I developed 10 essential LOs for delocalization and investigated how they are intended, enacted, and achieved. LOs are a teaching and learning tool as they provide guidance for educators (Kandlbinder, 2014) and are a metacognitive tool for learners (Towns, 2010). While a few studies have investigated delocalization (Taber, 2002; Betancourt-Pérez *et al.*, 2010; Xue and Stains, 2020; Brandfonbrener *et al.*, 2021), there has been no LOs that provide insight on what students should learn in a two-course organic chemistry sequence. This project answered the following research question: *How are the delocalization learning outcome being intended, enacted, and achieved in a two-course organic chemistry sequence?*

Overall, my research provides evidence for using specific learning and teaching tools to facilitate learners' construction of chemistry knowledge. In this thesis, I describe evidence-based ways to learn and teach about the symbolic and sub-microscopic level of chemistry.

## 1.1 References


Adbo K. and Taber K. S., (2009), Learners' Mental Models of the Particle Nature of Matter: A study of 16-year-old Swedish science students. *International Journal of Science Education*, **31**(6), 757–786.

American Educational Research Association, (2021), What is Education Research? https://www.aera.net/About-AERA/What-is-Education-Research

Balaban A. T. and Klein D. J., (2006), Is chemistry "The Central Science"? How are Different Sciences Related? Co-citations, Reductionism, Emergence, and Posets. *Scientometrics*, **69**(3), 615–637.

Betancourt-Pérez R., Olivera L. J., and Rodríguez J. E., (2010), Assessment of Organic Chemistry Students' Knowledge of Resonance-Related Structures. *Journal of Chemical Education*, **87**(5), 547–551.

Bodé N. E., Caron J., and Flynn A. B., (2016), Evaluating Students' Learning Gains and Experiences from Using Nomenclature101.com. *Chemistry Education Research and Practice*, **17**(4), 1156–1173.

Bongers A., Northoff G., and Flynn A. B., (2019), Working with Mental Models to Learn and Visualize a New Reaction Mechanism. *Chemistry Education Research and Practice*, **20**(3), 554–569.

Brandfonbrener P. B., Watts F. M., and Shultz G. V., (2021), Organic Chemistry Students' Written Descriptions and Explanations of Resonance and Its Influence on Reactivity. *Journal of Chemical Education*, **98**(11), 3431–3441.

Creswell J. W., (2012), Educational Research: Planning, Conducting, and Evaluating Quantitative, Fourth Eds. Prentice Hall Upper Saddle River, NJ.

Floden R. E., Wilson S. M., and Ferrini-Mundy J., (2001), Teacher Preparation Research: Current Knowledge, Gaps, and Recommendations: A Research Report Prepared for the US Department of Education and the Office for Educational Research and Improvement, Center for the Study of Teaching and Policy.

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. *Journal of Chemical Education*, **91**(11), 1855–1859.

Gilbert J. K., (2008), Visualization: An Emergent Field of Practice and Enquiry in Science Education, Gilbert J. K., Reiner M., and Nakleh M. (eds.). Visualization: Theory and Practice in Science Education. Models and Modeling in Science Education, vol 3. Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-5267-5_1

Henderson C., Connolly M., Dolan E. L., Finkelstein N., Franklin S., Malcom S., Rasmussen C., Redd K., and St. John K., (2017), Towards the STEM DBER Alliance: Why we Need a Discipline-Based STEM Education Research Community. *International Journal of STEM Education*, 4(1), 14.

Höffler T. N. and Leutner D., (2007), Instructional Animation Versus Static Pictures: A Meta-Analysis. *Learning and Instruction*, **17**(6), 722–738.

Johnstone A. H., (2000), Teaching of Chemistry - Logical or Psychological? *Chemistry Education: Research and Practice*, **1**(1), 9–15.

Johnstone A. H., (1991), Why is science difficult to learn? Things are seldom what they seem. *Journal of Computer Assisted Learning*, **7**, 75–83.

Johnstone A. H., (2010), You Can't Get There from Here. *Journal of Chemical Education*, **87**(1), 22–29.

Kandlbinder P., (2014), Constructive Alignment in University Teaching. *HERDSA News*, **36**(3), 5–6.

Kelly R. M., (2014), Using Variation Theory with Metacognitive Monitoring to Develop Insights into How Students Learn from Molecular Visualizations. *Journal of Chemical Education*, **91**(8), 1152–1161.

Kelly R. M. and Jones L. L., (2007), Exploring How Different Features of Animations of Sodium Chloride Dissolution Affect Students' Explanations. *Journal of Science Education and Technology*, **16**(5), 413–429.

Kozma R. B. and Russell J., (1997), Multimedia and Understanding: Expert and Novice Responses to Different Representations of Chemical Phenomena*. Journal of Research in Science Teaching*, **34**(9), 949–968.

National Research Council, (2015), How People Learn: Brain, Mind, Experience, and School: Expanded Edition, The National Academies Press.

Pazicni S. and Flynn A. B., (2019), Systems Thinking in Chemistry Education: Theoretical Challenges and Opportunities. *Journal of Chemical Education*, **96**(12), 2752–2763.

Schunk D. H., (2012), Learning Theories: An Educational Perspective, 6th ed. Pearson.

Stoyanovich C., Gandhi A., and Flynn A. B., (2015), Acid–Base Learning Outcomes for Students in an Introductory Organic Chemistry Course. *Journal of Chemical Education*, **92**(2), 220–229.

Taber K. S., (2002), Compounding Quanta: Probing the Frontiers of Students Understanding of Molecular Orbitals. *Chemistry Education Research and Practice*, **3**(2), 159–173.

Taber K. S., (2009), Learning at the Symbolic Level, in Multiple Representations in Chemical Education, Springers Netherlands, pp. 75–105.

Talanquer V., (2011), Macro, Submicro, and Symbolic: The Many Faces of the Chemistry "Triplet." *International Journal of Science Education,* **33**(2), 179–195.

Tasker R. and Dalton R., (2008), Visualizing the Molecular World - Design, Evaluation, and Use of Animations. *Visualization Theory and Practice in Science Education*, **3**, 103–131.

Towns M. H., (2010), Developing Learning Objectives and Assessment Plans at a Variety of Institutions: Examples and Case Studies. *Journal of Chemical Education*, **87**(1), 91–96.

Xue D. and Stains M., (2020), Exploring Students' Understanding of Resonance and Its Relationship to Instruction. *Journal of Chemical Education,* **97**(4), 894–902.

# Chapter 2. Theoretical frameworks

*"The theoretical framework is the structure that can hold or support a theory of a research study. The theoretical framework introduces and describes the theory that explains why the research problem under study exists." (Abend, 2008)*

In wet-lab chemistry, a chemist will choose a specific instrument or procedure for an experiment, and that instrument or procedure will affect the results. For example, to obtain information about the mass of a compound, a Mass Spectrometer (MS) would be used, as opposed to a Nuclear Magnetic Resonance (NMR). In social sciences, theoretical frameworks act as those instrument or procedure, they will guide the type of data obtain and what is appropriate to use.

A theoretical framework is the guiding lens to a study and consists existing theory used in a study. A theoretical framework consists of ideas, aims, goals, theories and assumptions about knowledge and research methodology (Crotty, 1998; Abend, 2008; Bodner and Orgill, 2008; Creswell, 2012). The framework helps to provide a theoretical basis for assumptions, connect the researcher to existing knowledge, generalize study finding within the theory, and identify limits to generalization.

My doctoral research was guided by several frameworks that I will introduce in this chapter. Those frameworks are

(1) modern information processing theory (IPT) (Chapter 5 and 6),

(2) mental models (Chapter 4),

(3) constructive alignment with backward design (Chapter 6 and 7), and

(4) causal reasoning (Chapter 7).

## 2.1 Modern information processing theory

The first theoretical framework is information processing theory (IPT) which is a cognitive theory (i.e., attempts to understand human thought processes) that describes how a person

learns new information. People perceive information through their senses and parse the new information in their working memory (WM) and then integrate the relevant information into the long-term memory (LTM).

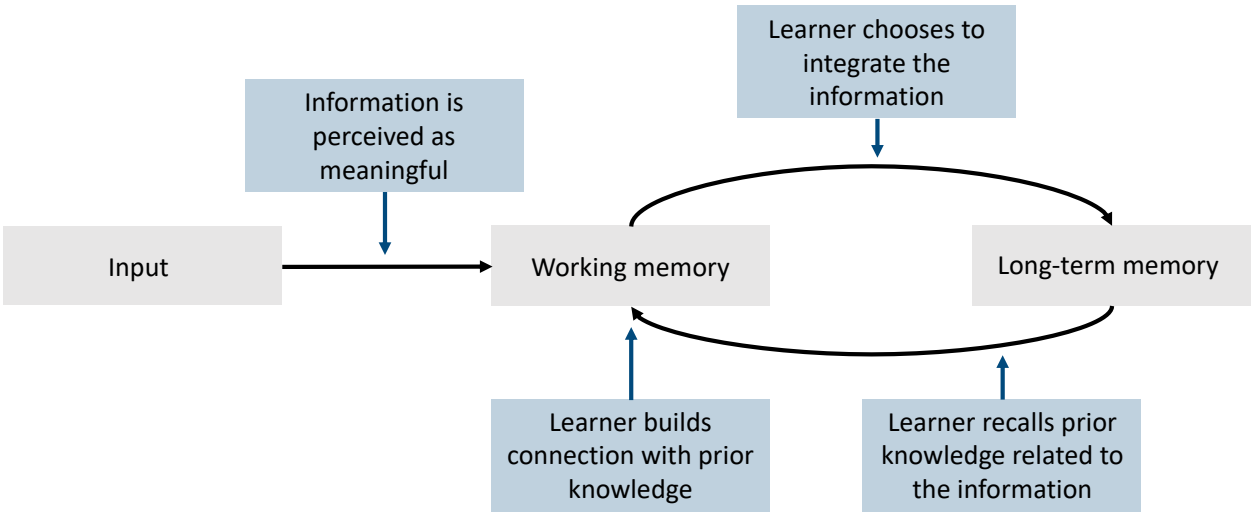



*Figure 2.1. Representation of the modern information processing theory*

The WM is the space where an individual manipulates and uses information (Baddeley and Hitch, 1974; Baddeley, 1992; Cowan, 1999; Chai *et al.*, 2018). A limited amount of information can be held and manipulated in the working memory at any given time (Miller, 1956; Luck and Vogel, 1997: Brady *et al.,* 2016). Long-term memory is where an individual store their knowledge that can be retrieved later. Information is stored as a network of concepts that are interconnected; recalling one concept can also activate the connected concepts. Information can be stored as declarative knowledge, which are facts and data obtained from events and experience, or procedural knowledge, which are the skills and procedures to perform a task (Cohen *et al.*, 1997; Gupta and Cohen, 2002).

The process described in IPT is akin to how computers would parse and file information. However, IPT fails to account that human have infinite computing power, and that people will be affected by their environment, motivation, and emotions. The theory was then adapted to account for those characteristics to yield the modern IPT. The modern information processing theory uses ideas from meaningful learning and constructivism (Piaget, 1977; Ausubel *et al.*,

1978, 1978; Vygotsky, 1978; Novak, 1993; Sweller, 1999; Kalyuga *et al.*, 2003; Mayer and Moreno, 2003; Schunk, 2016).

Meaningful learning theory, which supports IPT, says that for a learner to integrate knowledge, they must (1) perceive the information as meaningful, (2) link the new knowledge to prior knowledge, and (3) the learner must choose to integrate the knowledge (Ausubel et al., 1978; Novak, 1993). Therefore, for a new piece of information to be stored in the LTM, it should be meaningful for the learner, elaborate on prior knowledge, and be integrated in an organized network.

Constructivism, another theory that supports IPT, states that learners learn by constructing knowledge as opposed to passively taking in information (Piaget, 1977; Vygotsky, 1978; Novak, 1993). The theory explains that learners construct the knowledge from their experiences and promotes active learning, as opposed to passive teaching. While the instructors are facilitator and guide learners to build their knowledge, the onus of learning resides in the learner. The learner must be actively involved to elaborate the knowledge.

Storage problems will occur when a learner has no prior knowledge to connect the new information from, rendering it meaningless. Conceptually meaningless information can be stored in the LTM; however, it may not be part of a network and may not be retrieved. Another problem with storage is that the information may not be elaborated properly and fail to create the connection to other concepts. Therefore, links between concepts should be explicitly mentioned when teaching, despite seeming obvious to an instructor.

## 2.2 Mental models

The second theoretical framework used for this thesis (Chapter 4) is mental models. Models are fundamental tools for understanding sciences and can be distinguished as: conceptual models and mental models. Conceptual models are scientifically agreed models that is used by the scientific community to communicate. These models have been studied by scientist and have been agreed upon using the evidence obtained. On the other hand, mental models are internalized representations that are ever changing and abstract (Greca and Moreira, 2000). Mental models have been defined as “Internalized, organized knowledge structures that are used

to solve problems. They are encoded with respect to the spatial, temporal, and causal relationship of a concept" (Rapp, 2005).

Since mental models are internalized knowledge, they are difficult to conceptualize and define (Jones, et.al., 2011). Two prevalent school of thoughts are that a mental model is (1) a representation of an experience (Johnson-Laird, 1983), or (2) a characterisation of knowledge and cognitive processes (Gentner & Stevens, 1983). Mental models can also be conceptualized depending on which memory hold the mental model and how permanent they are. Some researchers claim mental models are (1) short-term items in the working memory (Johnson-Laird, 1983), (2) knowledge structures in the long-term memory (Craik, 1943), or (3) both (Nersessian, 2002). Despite differing perspectives, mental models are a good tool to investigate learning and reasoning (Jones et.al., 2011). Within my PhD work, I will be using the following conceptualization of mental models.

A person will build mental models around aspects of the physical and social world they know (Bower and Morrow, 1990). These models are fluid, meaning that they can change over time and are manipulated when a person thinks, plans, and explains phenomena. An individual's mental models are not right or wrong – they are a representation of that individual's experiences. New experiences, new information, or new situations will often result in changes to an individual's mental models.

Mental models have three main characteristics: (1) they represent a set of possibilities, (2) they are iconic, and (3) they represent what is true at the expense of what is false. The first characteristic means that an individual will have multiple mental models that represent different possibilities.

When faced with a specific situation, a person will use a working mental model, which are contextualized mental models that are temporal and contain only the necessary elements (Johnson-Laird, 2010). Individuals will construct a working mental model from their own mental models that they have constructed from previous experience. The working mental model is temporal and linked to the context in which it is used. Problem solving (i.e., finding solutions to a specific problem) and reasoning (i.e., thinking logically about something) are related to mental

models (Greca and Moreira, 2000; Johnson-Laird, 2001, 2010; Khemlani et al., 2014). Faulty reasoning can occur when mental models are not properly retrieved or if mental models are not in agreement with the scientifically agreed conceptual model. Mental models are key for problem-solving capabilities and for learning about the sub- microscopic level of chemistry (Bodner and Domin, 2000). Students form their own mental models about chemical reactivity and properties which they can use to reason about chemistry.

People can hold multiple mental models of the same concept in their long-term memory (Nersession, 2002). Those mental models are built from experiences and interactions with the concept (Bower and Morrow, 1990). When faced with a concept, one or some specific mental models that are relevant to the prompt/context at hand (Figure 2.2) will be used to solve a problem or reason about something. The problem and context, along with the mental models of the concepts form the temporal working mental model.

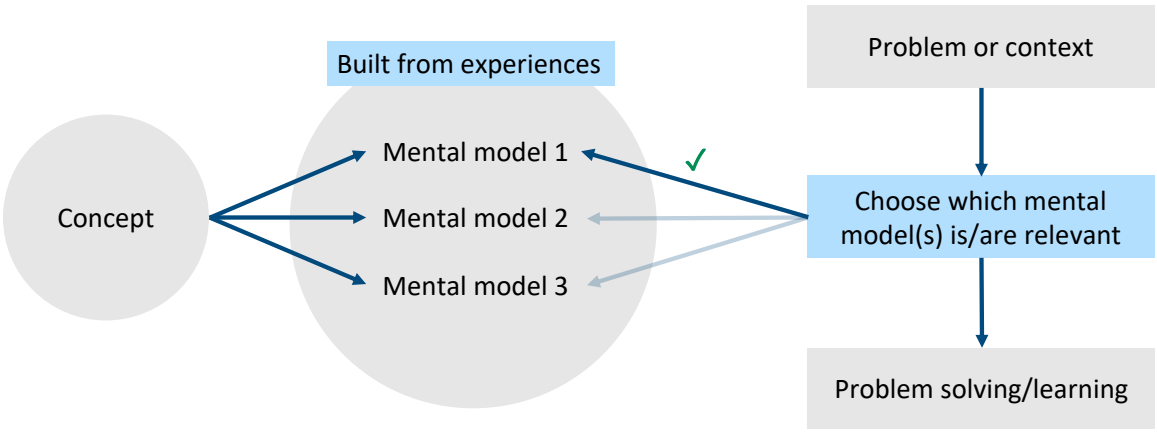



*Figure 2.2. Schematic representation of how people build and use mental models*

## 2.3 Constructive alignment framework

Constructive alignment is a framework for designing learning environment that focuses on what the learners need, as opposed to what the teachers does (Biggs and Tang, 2011). In constructive alignment, the learning outcomes (LOs), assessment, pedagogy, and other instructional decisions are carefully aligned (Figure 2.3)(Biggs and Tang, 2011; Ralph et al., 2022).

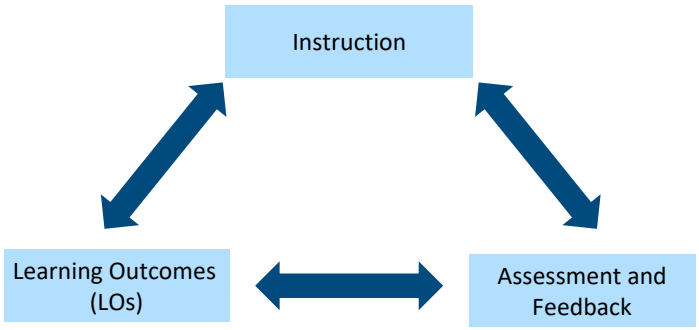



*Figure 2.3. Constructive alignment*

Key to constructive alignment are LOs, which are determined prior to instruction and assessment (Biggs and Tang, 2011). LOs are statements that describe the skill and knowledge a learner will have after instruction (Krathwohl, 2002; Brabrand and Dahl, 2009; Elmgren et al., 2015). A well worded LO contains an action verb that describes skills, a description of the content/knowledge to be learnt, and the context of how the LOs should be achieved (Figure 2.4)(Carey et al., 2001).

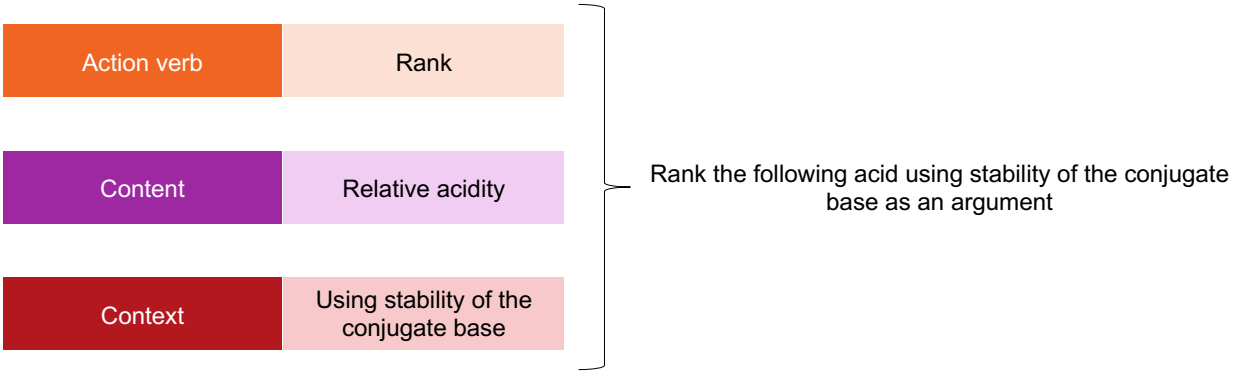



*Figure 2.4. Format of a LO*

LOs can be used both to guide teaching and assessment, but also to communicate expectations to learners, and as a metacognitive tool (Biggs, 1988; Towns, 2010; O'Connor et al., 2021). Educators can use the LOs to guide teaching and assessment, but also communicate expectations to learners. The LOs can be communicated in the course syllabus, at the beginning and end of a module or course, before assessments, or at other points in the instruction period. The learners can use the LOs as a tool to determine if they have achieved the expected skills.

LOs can be separated into three categories: intended, enacted, and achieved (Biggs, 1988; Bransford et al., 2000; Carle and Flynn, 2020; Raycroft and Flynn, 2020; Ralph et al., 2022). The

*intended* LOs are what the educator is hoping to achieve (i.e., what are students expected to learn?). The *enacted* LOs outcomes are how the LOs are being demonstrated and communicated to learners, which includes concepts like teaching, practice questions, assessments, the environment of the instruction. The *achieved* LOs represent the skills that the learners have demonstrated.

## 2.4 Reasoning

The fourth and final theoretical framework that guided my research is that of reasoning, which is thinking about something in a logical way to arrive at a conclusion. Reasoning has been the focus of many chemistry education research investigations, trying to elucidate how learners arrive at their conclusion. Several frameworks have been used to assess student reasoning in chemistry education, such as heuristics or Type I and Type II reasoning (McClary and Talanquer, 2011; Talanquer, 2017), abstractness and abstraction (Weinrich and Sevian, 2017), abductive reasoning (Wackerly, 2021), the rule-, case- and model- reasoning framework (Kraft et al., 2010), mechanistic reasoning (Sevian and Talanquer, 2014), among others.

A reasoning framework was developed by Sevian and Talanquer (2014) which was adapted from research in philosophy of science (Table 2.1) (Machamer *et al.*, 2000; Russ *et al.*, 2008). The framework has been used and expanded in several studies in chemistry education research (Moon *et al.*, 2016; Weinrich and Talanquer, 2016; Moreira *et al.*, 2018; Bodé *et al.*, 2019). The framework categorizes students' explanations or arguments in four levels, two of which have causality. Causal reasoning addresses the reasons WHY a phenomenon occurs (Talanquer, 2010; Sevian and Talanquer, 2014; Cooper, 2015; Weinrich and Talanquer, 2016). In chemistry, causal reasoning involves using underlying principles about molecules and atoms to explain a phenomenon and, therefore, the type of reasoning that learners are trying to learn and adopt. It implies a cause-and-effect relationship between components. In chemistry, causal reasoning involves using underlying principles about molecules and atoms to explain a phenomenon and, therefore, the type of reasoning that learners are trying to learn and adopt.

*Table 2.1. Description of the level of reasoning from Sevian and Talanquer (2014).*

| Causality | Level of reasoning | Descriptions |
|---|---|---|
| Non-causal | Descriptive | Describes the variable<br>Lacks causal arguments or relationship<br>No explanation why evidence is used |
| | Relational | Describes the variable<br>Provide a correlation between variables<br>Lacks causal arguments<br>No explanation why evidence is used |
| Causal | Linear | Describes the variable<br>Provide a correlation between variables<br>Uses causal arguments<br>Does not consider multiple causal arguments<br>Explains why evidence is used |
| | Multi-component | Describes the variable<br>Provide a correlation between variables<br>Uses causal arguments<br>Considers multiple causal arguments<br>Explains why evidence is used |

In *descriptive* reasoning, concepts are provided without including causality (Sevian and Talanquer, 2014). *Relational* reasoning involves outlining a relationship between two concepts; however, the underlying reason for that relationship is not explained (Sevian and Talanquer, 2014). Descriptive and relational reasoning are non-causal type of reasoning. In *linear causal* reasoning, the relationship between concepts is present and the reason is stated for why the concepts are important and how they relate to a conclusion (Sevian and Talanquer, 2014). Linear causal implies a line of reasoning that shows a cause and effect, for example statements like: "Something occurs BECAUSE of something". In *multi-component causal* reasoning, multiple linear causal relationships are involved. This type of reasoning involves weighing multiple factors and explaining why each is important, often involving an analysis of why one factor is dominant.

## 2.5 References


Abend G., (2008), The Meaning of 'Theory'. *Sociological Theory*, **26**(2), 173–199.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, (2014), *Standards for Educational and Psychological Testing*.

Ausubel D. P., Novak J. D., and Hanesian H., (1978), Educational psychology: A cognitive view, Rinehart and Winston (eds.) Holt.

Baddeley A., (1992), Working Memory. *Science*, **255**(5044), 556–559.

Baddeley A. D. and Hitch G., (1974), Working Memory, in Bower G. H. (ed.) Psychology of Learning and Motivation. Academic Press, pp. 47–89.

Biggs J. and Tang C., (2011), Aligning Assessment Tasks with Intended Learning Outcomes: Principles, in *Teaching for Quality Learning at University*, pp. 191–223.

Bodé N. E., Deng J. M., and Flynn A. B., (2019), Getting Past the Rules and to the WHY: Causal Mechanistic Arguments When Judging the Plausibility of Organic Reaction Mechanisms. *Journal of Chemical Education*, **96**(6), 1068-1082.

Bodner G. M. and Orgill M., (2008), *Theoretical Frameworks in Research in Chemistry and Science Education*, Pearson.

Bodner M. G. and Domin D. S., (2000), Mental Models: The Role of Representations in Problem Solving in Chemistry. *University Chemistry Education*, **4**(1), 24–30.

Bower G. H. and Morrow D. G., (1990), Mental Models in Narrative Comprehension. *Science*, **247**(4938), 44–48.

Brabrand C. and Dahl B., (2009), Using the SOLO taxonomy to analyze competence progression of university science curricula. High. Educ., 58(4), 531–549.

Brady T. F., Störmer V. S., Alvarez G. A., (2016), Working Memory is Not Fixed-Capacity: More Active Storage Capacity for Real-Woprld Object than for Simple Stimuli, *PNAS*, 113(**27**), 7459-7464

Bransford J. D., Brown A. L., and Cocking R. R., (2000), How people learn, Washington, DC: National Academy Press.

Carey L., Carey J., and Dick W., (2001), The Systematic Design of Instruction. 7th ed., *New Jersey Columbus, Ohio: Pearson*.

Carle M. S. and Flynn A. B., (2020), Essential learning outcomes for delocalization (resonance) concepts: How are they taught, practiced, and assessed in organic chemistry? *Chemistry Education Research and Practice*, **21**(2), 622–637.

Chai W. J., Abd Hamid A. I., and Abdullah J. M., (2018), Working Memory from the Psychological and Neurosciences Perspectives: A Review. *Frontiers in Psychology*, **9,** 401.

Cohen N. J., Poldrack R. A., and Eichenbaum H., (1997), Memory for Items and Memory for Relations in the Procedural/Declarative Memory Framework. *Memory*, **5**(1–2), 131–178.

Cooper M. M., (2015), Why Ask Why? *Journal of Chemical Education*, **92**(8), 1273–1279.

Cowan N., (1999), An Embedded-Processes Model of Working Memory. *Models of working memory: Mechanisms of active maintenance and executive control.*, 62–101.

Creswell J. W., (2012), *Educational research: Planning, conducting, and evaluating quantitative*, Fourth. Prentice Hall Upper Saddle River, NJ.

Creswell J. W. and Miller D. L., (2000), Determining Validity in Qualitative Inquiry. *Theory Into Practice*, **39**(3), 124–130.

Crotty M., (1998), *The foundation of Social Research: Meaning and Perspective in the Research Process*, SAGE Publication Ltd.

Elmgren M., Ho F., Åkesson E., Schmid S., and Towns M., (2015), Comparison and Evaluation of Learning Outcomes from an International Perspective: Development of a Best-Practice Process. *Journal of Chemical Education*, **92**(3), 427–432.

FitzPatrick B., (2019), Validity in Qualitative Health Education Research. *Currents in Pharmacy Teaching and Learning*, **11**(2), 211–217.

Greca I. M. and Moreira M. A., (2000), Mental Models, Conceptual Models, and Modelling. *International Journal of Science Education*, **22**(1), 1–11.

Gupta P. and Cohen N. J., (2002), Theoretical and Computational Analysis of Skill Learning, Repetition Priming, and Procedural Memory. *Psychological Review*, **109**(2), 401–448.

Johnson-Laird P. N., (2001), Mental Models and Deduction. *Trends in Cognitive Science,* **5**(10), 434–442.

Johnson-Laird P. N., (2010), Mental Models and Human Reasoning. *Proceedings of the National Academy of Sciences*, **107**(43), 18243–18250.

Kalyuga S., Ayres P., Chandler P., and Sweller J., (2003), Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist*, **38**, 23–31.

Khemlani S. S., Barbey A. K., and Johnson-Laird P. N., (2014), Causal Reasoning with Mental Models. *Frontiers in Human Neuroscience*, **8**(October), 1–15.

Kraft A., Strickland A. M., and Bhattacharyya G., (2010), Reasonable Reasoning: Multi-Variate Problem-Solving in Organic Chemistry. *Chemistry Education Research and Practice*, **11**(4), 281–292.

Krathwohl D. R., (2002), A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*, **41**(4), 212–218.

Luck S. J., Vogel E. J., (1997) The Capacity of Visual Working Memory for Features and Conjunctions. *Nature,* **390**, 279–281.

Machamer P., Darden L., and Craver C. F., (2000), Thinking about Mechanisms. *Philosophy of Science*, **67**(1), 1–25.

Maxwell J. A., (2012), *A realist approach for qualitative research*, Sage. Thousand Oaks, California

Mayer R. E. and Moreno R., (2003), Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist*, **1520**(38), 43–52.

McClary L. and Talanquer V., (2011), Heuristic Reasoning in Chemistry: Making Decisions about Acid Strength. *International Journal of Science Education*, **33**(10), 1433–1454.

Miller G. A., (1956), The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. *Psychological Review*, **63**(2), 81–97.

Moon A., Stanford C., Cole R., and Towns M., (2016), The Nature of Students' Chemical Reasoning Employed in Scientific Argumentation in Physical Chemistry. *Chemistry Education Research and Practice*, **17**(2), 353–364.

Moreira P., Marzabal A., and Talanquer V., (2019), Using a mechanistic framework to characterise chemistry students' reasoning in written explanations. *Chemistry Education Research and Practice*. **20**(1), 120-131

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

O'Connor E., Roy K., Walsh E., Huang D., Ke D. Y. J., Campbell Brown E., et al., (2021), The Evaluation of an Integrated Growth & Goals Module to Better Equip Students with Learning Skills in Postsecondary Courses: Systematic, Scalable, and Explicit. ChemRxiv

Piaget J., (1977), Psychology and Epistemology: Towards a Theory of Knowledge, Penguin Books.

Ralph V. R., Scharlott L. J., Schwarz C. E., Becker N. M., and Stowe R. L., (2022), Beyond Instructional Practices: Characterizing Learning Environments that Support Students in Explaining Chemical Phenomena. *Journal of Research in Science Teaching*, **55**(5), 841-875.

Rapp D. N., (2005), Mental models: Theoretical Issues for Visualizations in Science Education, in *Visualization in science education*, pp. 43–60.

Raycroft M. A. R. and Flynn A. B., (2020), What works? What's missing? An Evaluation Model for Science Curricula that Analyses Learning Outcomes through Five Lenses. *Chemistry Education Research and Practice*, **21**(4), 1110–1131.

Russ R. S., Scherr R. E., Hammer D., and Mikeska J., (2008), Recognizing Mechanistic Reasoning in Student Scientific Inquiry: A Framework for Discourse Analysis Developed from Philosophy of Science. *Science Education*, **92**(3), 499–525.

Saldaña J. and Omasta M., (2016), *Qualitative research: Analyzing life*, Sage Publications.

Schunk D., (2016), *Learning Theories: An Educational Perspective*, 6th ed. Pearson.

Sevian H. and Talanquer V., (2014), Rethinking Chemistry: A Learning Progression on Chemical Thinking. *Chemistry Education Research and Practice*, **15**, 10–23.

Sweller J., (1999), *Instructional design in technical areas*, Australia: ACER Press.

Talanquer V., (2017), Concept Inventories: Predicting the Wrong Answer May Boost Performance. *Journal of Chemical Education*, **94**(12), 1805–1810.

Talanquer V., (2010), Exploring Dominant Types of Explanations Built by General Chemistry Students. *International Journal of Science Education*, **32**(18), 2393–2412.

Towns M. H., (2010), Developing Learning Objectives and Assessment Plans at a Variety of Institutions: Examples and Case Studies. *Journal of Chemical Education*, **87**(1), 91–96.

Vygotsky L. S., Cole M., John-Steiner V., Scribner S., and Souberman E., (1978), Mind in Society: Development of Higher Psychological Processes, Harvard University Press.

Wackerly J. Wm., (2021), Abductive Reasoning in Organic Chemistry. *Journal of Chemical Education.*, **98**(9), 2746–2750.

Weinrich M. L. and Sevian H., (2017), Capturing Students' Abstraction while Solving Organic Reaction Mechanism Problems across a Semester. *Chemistry Education Research and Practice*, **18**(1), 169–190.

Weinrich M. L. and Talanquer V., (2016), Mapping Students' Modes of Reasoning When Thinking About Chemical Reactions Used to Make a Desired Product. *Chemistry Education Research and Practice*, **17**(394–406).

# Chapter 3. Methods

*"In planning a study, researchers need to think through the philosophical worldview assumption that they bring to the study, the research design that is related to this worldview and the specific methods and procedures of research translate into practice." – Creswell (2012, p.5)*

## 3.1 Overviews of methods in chemistry education research

The specific methods used in my work is explained in each chapter. However, in this chapter I explain the overarching methodological approaches I have taken in my research.

### 3.1.1 Ethic

Chemistry education aims to study the teaching and learning of chemistry and therefore our studies often involve human subjects. The Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (CIHR, NSERC, and SSHRC, 2018) guides the ethics for our research. The core principle of TCPS is:

*"Respect for human dignity requires that research involving humans be conducted in a manner that is sensitive to the inherent worth of all human beings and the respect and consideration that they are due. In this Policy, respect for human dignity is expressed through three core principles: Respect for Persons, Concern for Welfare, and Justice." (CIHR, NSERC, and SSHRC, 2108)*

The first consideration in a study should be the participants. Participants are volunteers who have the right to: informed consent, the ability to withdraw at any point, and to receive no harm from the study (Bauer, 2014; Taber, 2014; TCPS, 2018). In all my studies involving participants, the Research Ethics Board at the University of Ottawa granted permission for the research to proceed.

### 3.1.2 Type of research in education: qualitative and quantitative

Research in education can be separated into three broad categories: qualitative, quantitative, and mixed methods (Creswell, 2014). Mixed methods involve using a procedure that encompasses both quantitative and qualitative approaches.

Quantitative and qualitative research aim to answer different questions and have different purposes, sampling, and methods (Table 3.1). Qualitative research aims to understand the deeper meaning of a context and will rely on interpretative data while following a non-linear path (i.e., proceeds in circular back and forth) (Creswell, 2012; Neuman and Robson, 2014). Quantitative studies on the other hand rely on variables and hypotheses and precisely measuring variables in a linear fashion (Creswell, 2012; Neuman and Robson, 2014).

*Table 3.1. Differenced between qualitative and quantitative research*

| | Qualitative | Quantitative |
|---|---|---|
| Purpose | Explore: what, why, how | Providing evidence to a theory |
| Generalizable to a population | No | Yes (to specific population) |
| Depth | Very deep | Not as in deep |
| Number of participants | Few | Many |
| Type | Find meaning once immersed in the data | Testing a hypothesis posed by the researcher |
| Main idea | Concepts are themes | Concepts are distinct variables |
| Data types | Transcript, words, images, observations, etc. | Numbers |
| Theory production | Often inductive | Deductive |
| Methods | Particular to a specific study, replication is rare | Standard, with replication assumed |

### 3.1.3 Qualitative method – Thematic analysis

A theme provides a pattern to aspects of the data obtained. Themes do not need to capture the majority experience and are not about quantity, instead themes answer the research question in a meaningful way from the data (Ryan and Bernard, 2002; Braun and Clarke, 2006). A thematic analysis involves multiple steps to go from raw data (like a verbatim interview transcript) to themes (Figure 3.1) (Ryan and Bernard, 2002; Braun and Clarke, 2006; Creswell, 2014; Saldaña and Omasta, 2016; Charmaz and Henwood, 2017).

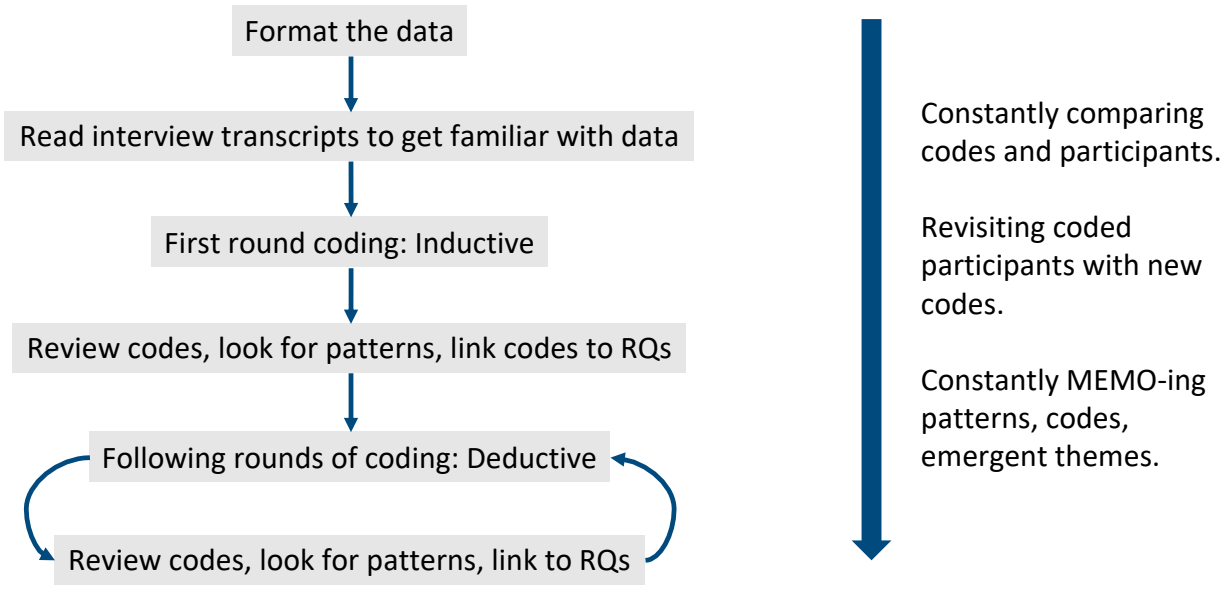



*Figure 3.1. General flow of thematic analysis*

The data obtained in qualitative research is usually extremely rich and in depth, and to analyze it the researcher will code the data. Coding is an analytical process of labelling the data to facilitate analysis and categorizing. Codes are units of data that can then be used to manipulate the information in the data (Creswell, 2014; Neuman and Robson, 2014; Saldaña and Omasta, 2016). For example, if a participant said the following "*the molecules are moving everywhere, right, left, up down, all over the place*" a code could be "Visualizing molecules as translating through space". There is no "right" way to code a piece of data – one person may assign a different code to the above quote (e.g., "Explaining the direction of movement") than someone else. Therefore, the research questions must guide the coding analysis.

Coding can be deductive or inductive (Creswell, 2012; Gibbs, 2018). Deductive coding means that the codes were decided before the analysis started, while inductive coding means the codes are formed from the data (Creswell, 2014; Saldaña and Omasta, 2016). Both types of coding are used in research settings, and typically multiple rounds of coding leading to more refined codes will occur. Coding in my research was done using gerund verbs to describe the actions and beliefs of the participants (Charmaz and Henwood, 2017) through the lenses of the research question.

When coding inductively, codes and data should be constantly compared and updated (i.e., re-coding previous interviews with new codes) (Corbin and Strauss, 2008). Through constant comparative analysis, categories are developed from the codes (Corbin and Strauss, 2008; Saldaña, 2013). The analysis can then move toward theme creation by narrowing the focus of the codes and categories (Ryan and Bernard, 2003).

### 3.1.4 Constructivist grounded theory: the role of the researcher

Constructivist grounded theory states that the conclusion emerges from the data and emphasizes the relationship between researcher and data (Novak, 1993; Bodner and Domin, 2000; Mills et al., 2007; Charmaz, 2008; McCall and Edwards, 2021). Constructivist grounded theory also understands that the researcher is not a "neutral observer" but has knowledge and perspective that are unique and will affect how the data is analyzed. In constructivist grounded theory, the researchers constantly refer to the research question, and they acknowledge and use their knowledge, skills, and experience to interpret the data (Dahlberg, 2006).

## 3.2 Positional statement: Potential biases in my work

Since bias comes can come from person experience, a positional statement gives information about the researcher's beliefs, perspectives, identities, etc., and provides the lens the researcher will use for analysis (Darwin Holmes, 2020). Since the analysis of my doctoral research was guided by my experiences and perspectives as well as the literature, a statement of position and a discussion of possible bias are necessary.

Over the years of my studies, I have gained substantial experience teaching, and my teaching philosophy is based on four core values. These values are: (1) aligning instruction, assessments, and expectations for students, (2) assessing students' progress, (3) developing an inclusive learning community, and (4) using reflective moments to improve my teaching. These beliefs about teaching may introduce biases in my research. While my values introduced biases, they also introduce experiences and expertise that guided my research.

I have a bachelor's degree in biochemistry and a master's degree in organic chemistry, both from the University of Waterloo. My experiences learning chemistry differed from the participants in my study due to my attending different universities. I also experienced my learning of organic chemistry in a different curriculum than the one used at the University of Ottawa, and I

have brought my perspective from how I learned into my research. In Chapter 6, I investigated the concept of delocalization, which is a concept I struggled with during my undergraduate studies. While there is evidence of learners' struggling with this concept (Taber, 2002; Betancourt-Pérez et al., 2010; Kim et al., 2019; Xue and Stains, 2020), I may have added some degree of personal bias in thinking – incorrectly – that everyone struggled the same way that I did (e.g., assuming that delocalization was not an important or useful concept and therefore ignoring it).

I started my doctoral research thinking that education research was like organic chemistry – with one clear answer for each problem. I was aiming to be as objective as possible; however, learning is not the same for everyone, therefore my research in education cannot be one size fit all. I have tried to use reflectivity and be explicit about my perspective to try to be as objective as possible. For example, during analysis I would always ask myself – Is this a theme because that is how I learned or is this a theme from the data?

### 3.2.1 Reliability and Validity

Since there is no correct, single understanding of the world, researcher's perspective and interpretations will affect how they collect, analyse and interpret at the data (Maxwell, 2012). Validity and reliability are two key components of education research that contributes to the trustworthiness of the findings. The two ideas aim to add rigour to the study by providing evidence that the findings are accurate, truthful, and consistent.

Reliability is approximately analogous to dependability, consistency, or precision. Reliability is making sure that the data obtained from an instrument or test would produce the same results in similar conditions (Neuman and Robson, 2014). For example, in organic chemistry, chemists rely on techniques (such as NMR) to measure their compounds. A reliable NMR instrument would provide the same results for the same compound and similar peaks in proton NMR spectra for similar protons. Chemist instinctually trust NMR spectrometers because the machine has been proven to be reliable. An example of reliability measure in education is a test-retest approach, where a person would take the same test at multiple times and the result should not vary (if no change is expected).

Validity indicates that we are testing what we mean to test (American Educational Research Association *et al.*, 2014). For example, in organic chemistry, if a chemist would like to know the exact mass of their compound, they would use a mass spectrometer (MS), as opposed to an NMR spectrometer. This is in a way what validity represents: making sure that the tools/method we are using are providing the right type of data.

The concept of validity is approximately analogous to accuracy. However, those concepts are not fully interchangeable. Accuracy refers to whether the measurement represents the "true" value of something. In educational research, there is not "true" correct answer, hence validity is used to determine if the measurements are aligned with what is desired to be measured.

Methods that are reliable will provide consistent results across different evaluators or conditions and the data collection and analysis is valid if it is appropriate for the goal of the experiment (Figure 3.2, left) (Neuman and Robson, 2014).


*Figure* 3.2. *Example of validity and reliability. Each red point represents a data point obtained that was aimed at the bullseye.*

Since no method will be perfectly valid and reliable all the time, we obtain evidence for validity on the data obtained rather than the method itself. The context and purpose of using the methods must be taken into consideration.

Validity in quantitative research has been established through the *Standards of Educational and Psychological testing* to include many types of different evidence for validity (American Educational Research Association *et al.*, 2014). The data obtained from instruments (i.e., tools that are used to assess a specific construct) will obtain several types of evidence of validity, such as test content, response process, internal structure, and relation to other variables

(Xu and Lewis, 2011; Arjoon *et al.*, 2013; American Educational Research Association *et al.*, 2014; Deng *et al.*, 2021). Evidence for reliability can be obtained via temporal stability, and internal consistency (Arjoon *et al.*, 2013; American Educational Research Association *et al.*, 2014). The types of validity and reliability in quantitative method is summarized in Figure 3.3. However, my thesis used mostly qualitative methods, and as such used a different set of validity and reliability measures.

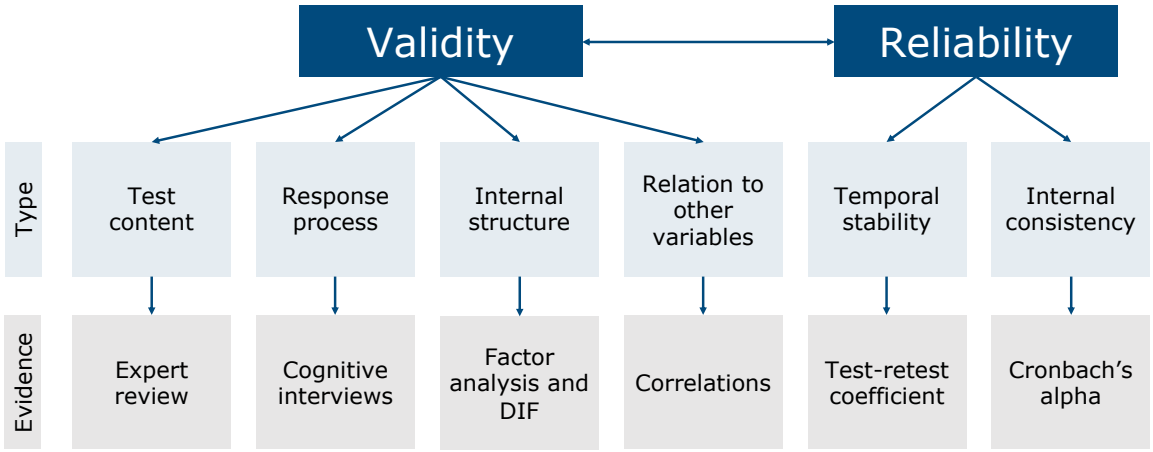



*Figure 3.3. Validity and reliability types along with the evidence required.*

## 3.3 Validity in qualitative research

Since my doctoral research is qualitative, I will focus on the validity measured related to this type of data. Evidence of validity in quantitative studies is used to show that a method assessed what it intended to assess within a specific population (American Educational Research Association et al., 2014). However, when dealing with rich qualitative data, validity refers to the inferences about the results; not validity of the data itself. Maxwell (2012) proposed several types of validity for qualitative data, with each providing validity to the different aspects of the analysis (Table 3.2). *Descriptive validity* refers to reporting what is visible, for example what the participants are doing. Providing description equips the reader with information about what has occurred and lets the reader decide on the credibility of the inferences the researcher has made. *Interpretative validity* is providing evidence that the interpretations of the researchers are accurate. In education qualitative research, we often report on thought processes or feelings, and the analysis should be backed by evidence that the researchers' interpretations are accurate.

*Theoretical* validity refers to evidence that the relationship, outlined by the researcher, between the data observed and the interpretation is accurate.

*Table 3.2. Types of validity in qualitative data as outlined by Maxwell (2012)*

| Type of validity | Description |
|---|---|
| Descriptive | Reporting physical and behavioral events that would in principle be observable |
| Interpretive | Validating real phenomena that are mental rather than physical. |
| Theoretical | Descriptive and interpretive understandings but takes these understandings a step further to develop an explanation of the relationships between concepts. |

To provide evidence for validity to qualitative data, several techniques can be used (Creswell and Miller, 2000; Saldaña and Omasta, 2016) which can be aligned with Maxwell (2012) types of validity.

*Table 3.3. Specific approach to obtain evidence for validity in qualitative research, aligned with Maxwell (2012) types of validity.*

| Specific approaches | Alignment with Maxwell's types of validity |
|---|---|
| Rich data (Creswell and Miller, 2000; Saldaña and Omasta, 2016) | Descriptive, Interpretive |
| Intensive, long-term involvement (Maxwell, 2012) | Descriptive, Interpretive |
| Thick description (Creswell and Miller, 2000) | Descriptive, Interpretive |
| Triangulation (Hesse-Biber and Leavy, 2010; Merriam and Tisdell, 2015) | Interpretive, Theoretical |
| Searching for discrepant evidence (Creswell and Miller, 2000) | Interpretive, Theoretical |
| Respondent validation (Maxwell, 2012) | Interpretive |
| Debriefing (Spall, 1998) | Interpretive |
| Analytic rigor (FitzPatrick, 2019) | Interpretive, Theoretical |

Collecting data that is rich in information and contains as much information as appropriate will help validate the findings (Saldaña and Omasta, 2016; FitzPatrick, 2019). Rich data gives the researcher the opportunity to have has much information as possible to infer their findings from. The data can come from the source itself, but also from field notes and

Long-term involvement implies that the research immerses themselves in the community. This provides the researcher with a personal insight of what they are trying to study.

Thick description involves describing in the research reports the actions, events, surrounding, what the participant is saying, body language etc. as necessary and acceptable.

Creswell and Miller (2000) explained that “credibility is established through the lens of readers who read a narrative account and are transported into a setting or situation”. To obtain this credibility thick, complete descriptions of what the participants said and did is important.

Using multiple sources of data helps strengthen the results by form of triangulation (Hesse-Biber and Leavy, 2010; Merriam and Tisdell, 2015). Looking for converging results from multiple data source adds validity. For example, looking for convergent results between and interview and a worksheet would provide some triangulation.

Looking for evidence within the data, that disproves the theory provides validity about the interpretation and theory that is generated (Creswell and Miller, 2000).

Respondent validity involves asking the participants to validate the interpretations made by the researcher (Maxwell, 2012). This type of evidence for validity is extremely important to make sure the researcher does not mis-interpret the study participants or attribute the participants something that is incorrect (Creswell and Miller, 2000).

While interpreting the data, the research can debrief with other researchers to validate their interpretations (Spall, 1998; FitzPatrick, 2019). Discussion between researcher will allow the researcher to question their interpretations and immerse themselves in the data. Debriefing can also provide support to a researcher and help them organize the data (Spall, 1998).

Finally, analytic rigor should be present in all aspect of the study, from data collection to analysis and reporting. Analytic rigor involves constantly circling from the data to the research questions to guide the research (FitzPatrick, 2019). While immerse in the study, documenting the findings, researchers’ standpoints, assumption is essential for the study to be rigorous (Saldaña and Omasta, 2016).

Within my research, we have used rich data, thick description, triangulation, searching for discrepant evidence, debriefing, and analytic rigor. My PhD research validity and reliability had covered all aspect (i.e., descriptive, interpretive, and theoretical).

### 3.3.1 Reliability in qualitative research

Evidence of reliability can be established in a variety of ways (Watts and Finkenstaedt-Quinn, 2021). To provide evidence of reliability two methods are commonly used: two coders

independently code a subset of the data, and one coder then codes the rest of the data, or, both coders code the entirety of the data and discuss to attain consensus (Campbell et al., 2013). So, other measures have been used that account for the potential of chance agreement – measures like Cohen's $\kappa$ (Cohen, 1960). Unfortunately, Cohen's $\kappa$ only works with two raters, does not account for missing data, and the interpretations are very case dependant (McHugh, 2012). Krippendorff $\alpha$ is another measure that is considered more robust since it calculates for agreements and disagreement, can accommodate multiple types of data (nominal, categorical, ordinal, and ratio), and can accommodate for missing data (Krippendorff, 2004; Watts and Finkenstaedt-Quinn, 2021). The generally accepted limit for Krippendorff $\alpha$ is a value over 0.7 (Krippendorff, 2004).

If the reliability criteria are not met, the two coders will meet and discuss any discrepancies. The code book will be updated to reach consensus. Once the code book has been updated, a second round of inter-rater reliability would be done that will hopefully have led to an acceptable agreement. This cycle may be done several times to obtain a codebook that lacks ambiguity and that can produce reliable results by whomever uses the method.

In my research, I asked a second person code 15% of the data. Our codes were then compared, and the percent agreement was calculated. The generally accepted limit is at least 80% agreement (McHugh, 2012). However, percent agreement does not account for chance agreement between codes (e.g., one of the coders simply guessing).

## 3.4 Methods used in my doctoral research

My doctoral research consists of four different studies. Each chapter will contain a detailed section about the specific method used for the study, including recruitment, participants, method, study design, and analysis.

## 3.5 References


American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, (2014), Standards for Educational and Psychological Testing, American Educational Research Association.

Arjoon J. A., Xu X., and Lewis J. E., (2013), Understanding the State of the Art for Measurement in Chemistry Education Research: Examining the Psychometric Evidence. *Journal of Chemical Education*, **90**(5), 536–545.

Bauer C. F., (2014), Ethical Treatment of the Human Participants in Chemistry Education Research, in *Tools of Chemistry Education Research*, ACS Symposium Series. American Chemical Society, pp. 279–297.

Betancourt-Pérez R., Olivera L. J., and Rodríguez J. E., (2010), Assessment of Organic Chemistry Students' Knowledge of Resonance-Related Structures. *Journal of Chemical Education*, **87**(5), 547–551.

Bodner G. M. and Domin D. S., (2000), Mental Models: The Role of Representations in Problem Solving in Chemistry. *University Chemistry Education*, **4**(1), 24–30.

Braun V. and Clarke V., (2006), Using Thematic Analysis in Psychology. Qualitative Research in Psychology, 3(2), 77–101.

Campbell J. L., Quincy C., Osserman J., and Pedersen O. K., (2013), Coding In-depth Semi-Structured Interviews: Problems of Unitization and Intercoder Reliability and Agreement. *Sociological Methods & Research*, **42**(3), 294–320.

Charmaz K., (2008), Constructionism and the Grounded Theory Method, in Handbook of constructionist research, Holstein J. A. and Gubrium J. F. (eds.)., Guilford Press.

Charmaz K. and Henwood K., (2017), Grounded Theory Methods for Qualitative Psychology, in The SAGE Handbook of Qualitative Research in Psychology, pp. 1–23.

Cohen J., (1960), A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, **20**(1), 37–46.

Creswell J. W., (2012), Educational research: Planning, conducting, and evaluating quantitative, Fourth ed. Prentice Hall Upper Saddle River, NJ.

Creswell J. W., (2014), Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, SAGE Publications.

Creswell J. W. and Miller D. L., (2000), Determining Validity in Qualitative Inquiry. *Theory Into Practice*, **39**(3), 124–130.

Dahlberg K., (2006), The Essence of Essences – the Search for Meaning Structures in Phenomenological Analysis of Lifeworld Phenomena. *International Journal of Qualitative Studies on Health and Well-being*, **1**(1), 11–19.

Darwin Holmes A. G., (2020), Researcher Positionality – A Consideration of Its Influence and Place in Qualitative Research – A New Researcher Guide. *International Journal of Education*, **8**(4), 1–10.

Deng J. M., Streja N., and Flynn A. B., (2021), Response Process Validity Evidence in Chemistry Education Research. *Journal of Chemical Education*, **98**(12), 3656–3666.

FitzPatrick B., (2019), Validity in Qualitative Health Education Research. *Currents in Pharmacy Teaching and Learning*, **11**(2), 211–217.

Gibbs G. R., (2018), Analyzing Qualitative Data, 2nd ed. SAGE Publication Ltd.

Hesse-Biber S. N. and Leavy P., (2010), The practice of qualitative research, Sage **16**(6), pp.455-468

Kim T., Wright L. K., and Miller K., (2019), An Examination of Students' Perceptions of the Kekulé Resonance Representation using a Perceptual Learning Theory Lens. *Chemistry Education Research and Practice*, **20**(4), 659–666.

Krippendorff K., (2004), Reliability in Content Analysis.: Some Common Misconceptions and Recommendations. *Human Communication Research*, **30**(3), 411–433.

Maxwell J. A., (2012), A realist approach for qualitative research, Sage.

McCall C. and Edwards C., (2021), New Perspectives for Implementing Grounded Theory. *Studies in Engineering Education*, **1**(2), 93.

McHugh M. L., (2012), Interrater Reliability: The Kappa Statistic. *Biochemia medica*, **22**(3), 276–282.

Merriam S. B. and Tisdell E. J., (2015), Qualitative Research: A guide to Design and Implementation, John Wiley & Sons.

Mills J., Chapman Y., Bonner A., and Francis K., (2007), Grounded Theory: A Methodological Spiral from Positivism to Postmodernism. *Journal of Advanced Nursing*, **58**(1), 72–79.

Neuman W. L. and Robson K., (2014), Basics of social research, Pearson Canada Toronto.

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

Ryan G. W. and Bernard H. R., (2002), Techniques to Identify Themes. *Field Methods*, **15**(1), 85–109.

Saldaña J. and Omasta M., (2016), Qualitative Research: Analyzing Life, Sage Publications.

Spall S., (1998), Peer Debriefing in Qualitative Research: Emerging Operational Models. *Qualitative Inquiry*, **4**(2), 280–292.

Taber K. S., (2002), Compounding Quanta: Probing the Frontiers of Students Understanding of Molecular Orbitals. *Chemistry Education Research and Practice*, **3**(2), 159–173.

Taber K. S., (2014), Ethical Considerations of Chemistry Education Research Involving 'Human Subjects.' *Chemistry Education Research and Practice*, **15**, 109–113.

Watts F. M. and Finkenstaedt-Quinn S. A., (2021), The Current State of Methods for Establishing Reliability in Qualitative Chemistry Education Research Articles. *Chemistry Education Research and Practice*, **22**, 565–578.

Xu X. and Lewis J. E., (2011), Refinement of a Chemistry Attitude Measure for College Students. *Journal of Chemical Education*, **88**(5), 561–568.

Xue D. and Stains M., (2020), Exploring Students' Understanding of Resonance and Its Relationship to Instruction. *Journal of Chemical Education*, **97**(4), 894–902.

# Chapter 4. Molecular dynamic visualization: what are participants mental models of the sub-microscopic level?

*"Using models and developing mental models are invaluable skills for students to learn, and our findings imply that organic chemistry animations can be used to bolster students' mental models to include dynamic interactions of molecules and particles." (Bongers, Beauvoir, Streja, Northoff and Flynn, 2020)*

## 4.1 Problem and context: Students' mental model of the sub-microscopic level may not be scientifically accurate

A cohesive understanding of chemistry relies on building an internal representation (i.e., mental model) of the sub-microscopic level (Johnstone, 1991; Kozma and Russell, 1997; Talanquer, 2011). The sub-microscopic level is highly dynamic; chemical processes are random interactions and collisions between particles. Visualizing this dynamic and random environment can sometimes be difficult for students (Adbo and Taber, 2009; Strickland et al., 2010; Yan and Talanquer, 2015; Popova and Bretz, 2018). In this chapter, I explored the mental models of learners about four aspects: (1) motion, (2) collisions, (3) amounts and proportions, and (4) probabilistic thinking. These aspects are hard to depict in a static image and can be shown as a series of time points. Using static images requires learners to imagine what molecules are doing. Dynamic visualizations can show these concepts so that learners do not have to fill in the blanks.

## 4.2 Literature review: Mental models of the sub-microscopic level

Representations have been defined as: "*a likeness or simulation of some idea, concept, or object*" (Rapp and Kurby, 2008). A representation can be external, (i.e., something physical to make an abstract concept visible), or internal, (i.e. a representation that is constructed mentally by a person) (Figure 4.1) (Rapp, 2005; Rapp and Kurby, 2008; Uttal and O'Dogerty, 2008). Visualizations are external representations that are visual, including graphs, schemes, videos, molecular models, etc. Visualizations can be used by a person to build an internal representation of a concept (Gilbert, 2008).

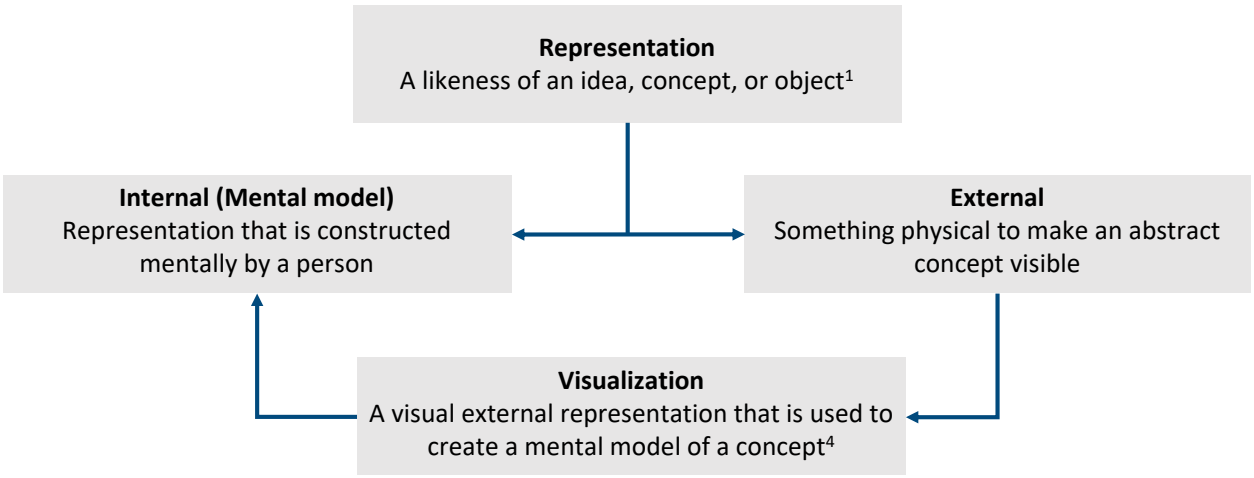



*Figure 4.1. Graphical relationships of the type of representation*

While physical models and pictures have been commonplace due to low cost and ease of use, they lack the dynamic, collisions, and probabilistic nature of the submicroscopic world. Technological advances have given educators the tools of animation and simulation that show a more accurate visualization of the submicroscopic domain. Dynamic visualizations have been shown to lead to higher conceptual and procedural learning gains than static images (Höffler and Leutner, 2007). Using dynamic visualization can help learners visualize the sub-microscopic level (Stieff and Wilensky, 2003; Velázquez-Marcano et al., 2004; Tasker and Dalton, 2006; Kelly and Jones, 2008; Kelly and Akaygun, 2016; Kelly et al., 2017, 2017, 2021). Four concepts that are hard to depict in static images are: (1) motion, (2) collisions, (3) amounts and proportions, and (4) probabilistic thinking. These four concepts formed the basis of this research, which I conducted as part of a team.

### *4.2.1* Particles are dynamic

Particles are constantly moving internally (i.e., bond vibration and rotations) and in space (i.e., displacement). Unfortunately, some learners may have the alternate conception that molecules and atoms are static (Tasker and Dalton, 2008). Some students have been shown to treat symbolic representations as static entities (Bhattacharyya and Bodner, 2005; Ferguson and Bodner, 2008; Adbo and Taber, 2009; Strickland et al., 2010). However, after viewing computer animations that show the dynamic motion of molecules students tend to have a more dynamic mental model of the sub-microscopic level (Tasker and Dalton, 2006; Sanger et al., 2007; Tasker and Dalton, 2008; Venkataraman, 2009; Akaygun and Jones, 2013; Tang and Abraham, 2016).

Some students also believe that atoms in solids were not moving and that they were stuck together (Adbo and Taber, 2009).

### 4.2.2 There are many particles in present in a solution

Learners often harbor alternate conceptions about the proportion and amount of matter. In previous work, learners often only focused on the reacting molecules and the product, without paying attention to solvents or other molecules when talking about reaction coordinate diagrams (Popova and Bretz, 2018). However, opposing results were found in a different study, most learners showed multiple of each species when asked to draw a storyboard of dissolution (Kelly and Jones, 2007a). This shows that students either chose to focus on only the reactants or hold the conception that only one molecule will react. The proportion of molecules present in a solution can also be difficult for learners to conceptualize (Rosenthal and Sanger, 2012a; Kelly and Akaygun, 2016). Other studies have found that visualizations did not remedy alternate conceptions and sometimes added more alternate conceptions (Kelly and Jones, 2007a; Akaygun and Jones, 2013). Some learners showed that a single water molecule would carry ions away from the lattice, while most of the learners simply omitted the role of the water molecules (Kelly and Jones, 2007a; Kelly and Akaygun, 2016), or dismiss it as the solvent (Rosenthal and Sanger, 2012a). Organic chemistry is often taught showing a mechanism of a reaction, which depicts a single molecule of the reacting species, and often does not mention solvents. This way of teaching may lead learners into thinking there are few molecules present.

### 4.2.3 Reactions occur through collisions

For a reaction to occur, two particles must collide in the correct orientation with enough energy. In investigations of students' skills in interpreting reaction coordinate diagrams (i.e., static, symbolic level) participants rarely mentioned motion or collisions (Popova and Bretz, 2018; Atkinson et al., 2021). However, other work about energy and pressure found that most participants mentioned collisions (Macrie-Shuck and Talanquer, 2020). Research has shown that participants gained a better understanding of collision theory after viewing an animation (Sweeder et al., 2019). However, the contexts of each study were very different. The participants in the Popova and Bretz (2018) study were provided with a symbolic representation of a reaction coordinate diagram and the Lewis structure of plausible mechanisms, while the participants in

the Macrie-Shuck and Talanquer (2020) study were provided with particulate drawings. The Sweeder et al. (2019) study used written answers (with some particulate drawings) along with a simulation. From the literature, we can hypothesize that when the context is mostly symbolic, participants did not use language that involve collisions. However, when participants are faced with a sub-microscopic depiction, they mention collisions.

### 4.2.4 The sub-microscopic level is random and probabilistic

Probabilistic thinking in chemistry is thinking that reactions and motion in the sub-microscopic level are random and that collisions are probabilities, which can be difficult to understand for learners (Garvin-Doxas and Klymkowsky, 2008; Jolfaee et al., 2014; Savard, 2014) (Figure 4.2). Due to the challenges associated with probabilistic thinking, learners tend to show deterministic thinking about reactions. Learners who use a deterministic thinking approach tend to explain the reason why a reaction as

- external forces drive the reaction (Yan and Talanquer, 2015),
- something causes the reaction (Taber and García-Franco, 2010; Yan and Talanquer, 2015),
- because there is an active (i.e., reactive) species (Taber and García-Franco, 2010; Yan and Talanquer, 2015),
- molecules want something or have goals (Talanquer, 2007, 2013; Yan and Talanquer, 2015),

Despite learners adopting deterministic thinking mental models, the currently accepted mental model is probabilistic where only probabilities and randomness affect the sub-microscopic level.

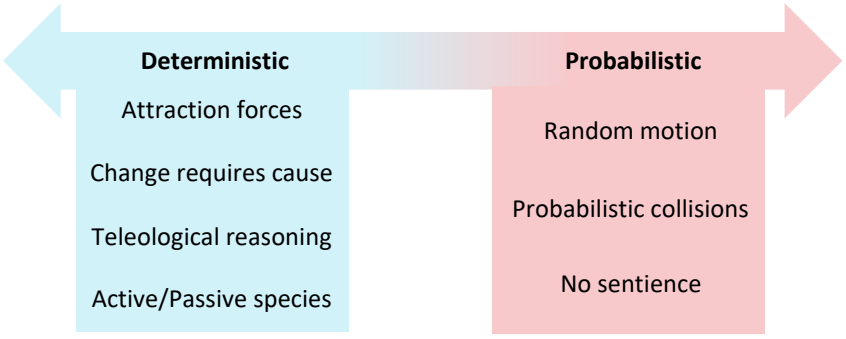



*Figure 4.2. Deterministic and probabilistic mental models*

Collisions are Random, but motion of molecules is also random. Some students interviewed by Adbo and Taber (2009) mentioned that some erroneous factors (i.e., macroscopic

input of heat) could affect the motion of particles. This study interviewed 16-year-old students and it may not be a belief that is held by more advanced students. Macrie-Shuck and Talanquer (2020) showed that many students could identify that intermolecular forces and energy act on a particle, and none of the students mentioned that those forces would affect particle speed or direction (Macrie-Shuck and Talanquer, 2020b).

The literature has shown that most learners have a deterministic mental model of the submicroscopic level. I hypothesize that most learners will not show a probabilistic mental model due to the way we currently teach chemistry.

## 4.3 Literature Review: Molecular dynamic visualization as a tool for learning

A rich mental model of chemistry includes the sub-microscopic level, symbolic level, and macroscopic level (Kozma and Russell, 1997; Taber, 2009; Talanquer, 2011). A robust mental model of the sub-microscopic level provides a more cohesive understanding of chemistry since it provides the learners with the tool to understand why and how some phenomena occur. Tasker and Dalton (2006) reported that learners may not have a rich mental model of the sub-microscopic level (Tasker and Dalton, 2006) and therefore may not understand why or how reactions occur (Yan and Talanquer, 2015). Learning tools that can help learners build dynamic, probabilistic mental models are therefore needed to demonstrate how reactions occur. Visualizations have the potential to help students build a robust mental model about chemistry and scientific accuracy learners' mental models (Suits, 2015).

There have been many reports that visualizations help learners build scientifically accurate mental models (Williamson and Abraham, 1995; Rosenthal and Sanger, 2012a, 2012a; Jones and Kelly, 2015; Akaygun, 2016; Kelly *et al.*, 2017; Bongers *et al.*, 2019; Kelly *et al.*, 2021) show that visualizations led to lasting benefits (Ardac and Akaygun, 2004; Tasker and Dalton, 2006; Abraham *et al.*, 2010). The benefits included learner's using particulate drawing when explaining a chemical phenomenon 15 months later (Ardac and Akaygun, 2004) and higher scores on the in-course exam that was administered two weeks after viewing the visualization (Abraham *et al.*, 2010).

## 4.4 Molecular dynamic visualizations

Several frameworks have been proposed for how learners gain information from visualization (Notoel, et.al., 2022), including Mayer's cognitive theory of multimedia learning (Moreno and Mayer, 2007; Mayer, 2005; 2014). The cognitive theory of multimedia learning (Mayer & Chandler, 2001) takes roots in Baddeley's model of working memory (Baddeley, 1983, 2001) and Pavio's dual-coding theory (Clark and Pavio, 1991). The theory states that learners have two separate channels for processing visual and auditory information, which both have limited capabilities. The learner must integrate the information into pre-existing knowledge within their working memory, which is limited (Baddeley, 1983; Sweller, 1994). The cognitive theory of multi-media learning is related to meaningful learning (Novak, 1993), in that a learner needs to (1) perceive the information as meaningful, (2) link the new knowledge to prior knowledge, and (3) the learner must choose to integrate the knowledge (Mayer, 2005; 2014). According to the cognitive theory of multi-media learning theory, cognitive load and learners' prior knowledge are key components to consider while designing visualizations.

### 4.4.1 Cognitive load

A person's working memory can only hold about seven pieces of information (Miller, 1956) and the amount of information is referred to as the cognitive load. Some information can be chunked together into a larger piece of information, allowing more information to be stored in working memory (Schunk, 2016). There are three types of cognitive load that a learner will encounter: intrinsic , extraneous, and germane. Intrinsic cognitive load occurs inside the learner's mind and involves the obtaining, organizing, and linking of new knowledge to prior knowledge. Germane cognitive load is the obtaining and storing of information into the long-term memory. Extraneous cognitive load is information that is unnecessary for the learner so it should be reduced as much as possible. Visualization will aim to provide intrinsic load by helping the learners build the information, extraneous load needs to be reduced so the learner can focus on important information.

The cognitive load in a molecular dynamic visualization will typically be high since learners must interpret the information, make sense of what they are seeing, and create or update mental models with the information (Mayer, 2005). In fact, dynamic visualizations can lead to cognitive

overload because there is so much information being transmitted. When cognitive overload occurs, a visualization will have a similar effect on learning compared to static images (Tversky et al., 2002; Höffler and Leutner, 2007). However, visualizations have the potential to both help reduce cognitive load by (1) avoiding decorative and useless information, (2) focusing on the critical components, and (2) be clear about the conceptual knowledge to be learnt (Tversky et al., 2002; Tasker and Dalton, 2006).For example, during a study about molecular symmetry learners mentioned that "I no longer need to imagine the structure in my mind" and that they could focus on learning chemistry (Tuvi-Arad and Gorsky, 2007; Venkataraman, 2009).

Several studies have shown that cognitive load does affect how learners use visualization. Learners preferred and gained more from visualizations when viewing simpler animations before complex ones (Rosenthal and Sanger, 2012, 2013), or proceeding from something familiar to something new (Williamson et al., 2012). Students who viewed a simpler animation before a complex animation (of the same process) recognized more cues in the complex animation, and their explanations were more scientifically accurate (Rosenthal, 2013). When viewing the complex animation first, the participants would state becoming overwhelmed (Rosenthal and Sanger, 2012, 2013). When viewing the simple animation first, participants were better at recognizing molecular elements of the more complex animation (Rosenthal and Sanger, 2013). Two reasons have been proposed for this effect; first, that the simpler visualization would cue participants to the icons and symbols in the more complex animation and second, that the simpler animation had fewer components on the screen therefore reducing the extraneous cognitive load for learners. Similarly, viewing a macroscopic level animation before viewing a molecular animation seemed beneficial to learners because the macroscopic video could act as a cue for the molecular animation (Williamson et al., 2012).

### 4.4.2 Learners' prior knowledge

People with different prior knowledge are likely to focus on different parts of the animations (Tasker and Dalton, 2006; Kelly et al., 2017). Novices tend to focus on surface features (e.g., specific reaction), while experts typically explore the deeper meaning (e.g., broader terms of reactivity) (Günersel and Fleming, 2013; Kelly and Akaygun, 2019). Learners with a lower prior knowledge gained more from a molecular dynamic visualization than learners with higher prior

knowledge did (Tasker and Dalton, 2006; Ollino et al., 2018). Since mental models are built from experiences (Bower and Morrow, 1990), experts will have more experience extracting information from a visualize than learners. Viewing different animations about a similar process can expend a learner's mental model by showing variation. Participants in one study were more successful in predicting reactions after viewing three different visualizations, as opposed to just one or two (Velàzquez-Marcano et al., 2004). Similarly, in a different study, learners attended to features that appeared in two animations that otherwise varied in proposed mechanism (Kelly et al., 2017). In the study by Kelly et al., learners found that a feature present in two animations must be important and scientifically accurate.

People gain information from visualization in different ways, which are dependent on several factors. Kelly (2014) outlined five ways learners used the animation to change their pictorial representation after viewing the animation: (1) as a verification source (i.e., using the visualization as evidence to support their drawing), (2) recollection trigger (i.e., reminder of forgotten information), (3) imitator (i.e., replicating the visualization without being able to explain the concepts), (4) short-cuts in constructing knowledge, and (5) misinterpreting the visualization. Therefore, learners will react and learn differently from visualizations with some using the visualization to improve their work, while others will simply imitate or disregard the visualization (Kelly, 2014).

Visualization can be used to correct learners' alternate conceptions by showing them the scientifically accepted conception (Stieff and Wilensky, 2003) but visualizations also have the potential of introducing new alternate conceptions. For example, some learners drew liquid water with molecules further apart (like in the gas phase) after viewing an animation (Kelly and Jones, 2007). Another alternate conception that learners obtained from animation was that water had no role in the dissolution of salt, or that if it did, a single water molecule would carry the ion away (Kelly and Jones, 2007). The literature shows that when gaining information from visualizations novices may misinterpret the visualization. Therefore, visualizations need to be careful of how the information is presented, such as avoiding extraneous (i.e., useless) information (Mayer, 2005; Mayer and Estrella, 2014), and using various technique to focus on key criteria (Bussey and Orgill, 2015; Suits, 2015).

Some research suggests that learners may show the correct conceptualization on paper but fail to explain it properly (Ardac and Akaygun, 2004) or will draw what they saw in animation, without being able to explain the underlying concept (Kelly and Jones, 2007).

Balancing scientific accuracy, cognitive load and learners' prior knowledge can be difficult when designing animations. Scientific accuracy… Reducing cognitive load can be achieved by giving the learner autonomy over the pace and content of the animation (José and Williamson, 2005). Pre-training (e.g., explaining the representations, providing background on the visualization) can provide learners with the prior knowledge needed to understand the visualization (Moreno and Mayer, 2007; Mayer, 2005; 2014). Visualizations can be used to correct learners' alternate conceptions by showing them the scientifically accepted conception (Stieff and Wilensky, 2005; Tasker and Dalton, 2006, 2008). Therefore, for our research we decided to use Odyssey (Wavefunction, 2019), a scientifically accurate simulation that provides user autonomy and pre-training.

## 4.5 Learning tool: Odyssey by Wavefunction

Several visualizations in chemistry exist and have been researched such as PhET (Akaygun and Jones, 2013; Moore et al., 2013, 2014; Watson et al., 2020), VisChem (Tasker and Dalton, 2006, 2008; Yezierski and Birk, 2006; Kelly and Jones, 2007), Organic ChemWare (Bongers et al., 2019, 2020). Our research focused on the simulation created within the Odyssey (Wavefunction, 2019) program. We selected two simulations that demonstrated the four aspects we planned to study (i.e., motion, collisions, amounts, and randomness). The first simulation showed a solution of acetone and water and aimed to show students how many molecules are present in a solution, as well as how they move randomly. No reaction or collisions were visible in the simulation. The second simulation was a gas phase reaction between fluorine gas ($F_2$) and nitrogen dioxide ($NO_2$). This simulation aimed to show learners successful and unsuccessful collisions, as well as the random motion of molecules.

## 4.6 Goals and research questions

The goal of my research was to explore participants' mental models of the sub-microscopic level, focusing on motion, amounts, collisions, and probabilistic thinking. We are not

probing for the mental models that participants hold but instead exploring which mental models the participant chose to use. Therefore, the research questions guiding this work are:

**RQ1**: What are participants' mental models of the sub-microscopic level?

**RQ2**: What mental models are participants using before and after using a molecular dynamic visualization?

**RQ3**: How are participants using the visualization to learn?

## 4.7 Manuscript in preparation: Investigating participants mental models of the sub-microscopic level

*Exploring students use of dynamic, collision-based, probabilistic mental models: influence of context and conflicting mental models*

Myriam S. Carle, Peter G. Mahaffy, and Alison B. Flynn

**Abstract:** Chemical processes are dynamic and random; however, students tend to have a static and non-random view of the sub-microscopic level. Molecular dynamic simulations can help learners build a dynamic mental model by showing the dynamic aspect of the submicroscopic environment. We administered two worksheets and two interviews to participants (N = 9) both before and after they viewed a molecular dynamic visualization. The four main findings are (1) all participants used a dynamic motion mental models and a collision-based mental model, (2) participants also used multiple mental models of the sub-microscopic level and use the appropriate mental model for a problem or context, (3) the majority of participants (N =7) demonstrated conflicting mental models about the randomness of the sub-microscopic level, and (4) eight participants showed cognitive dissonance (i.e., dismissing the simulation, interpreting the simulation to match their previously held mental model, or using the simulation to change their mental models) when watching the simulation. The findings have implications that simulations can be a great tool to demonstrate random collisions particle in motion mental models, and that they can create cognitive dissonance that leads to new mental models being built. Our work also provides the implication that dynamic mental models can be elicited from students depending on the context.

**Author Contribution:** MSC and ABF brainstorm the idea and design of the study. MSC did the data collection, data analysis, and manuscript write up. ABF provided guidance and feedback throughout the entirety of the project.

## 4.8 Introduction

Chemistry encompasses every substance and affects everything in our everyday life, but it can be difficult to learn because many aspects of chemistry are invisible to the human eye (Johnstone, 1991). Understanding what is occurring at a scale we cannot see is difficult for learners to conceptualize (Kozma and Russell, 1997). The sub-microscopic level is a highly dynamic environment in which molecules are constantly moving and colliding. Being able to visualize the complex and dynamic processes at the sub-microscopic level can help people build a robust mental model of both chemistry and chemical reactivity. However, previous reports have showed that students: treat symbolic representations as static entities (Adbo and Taber, 2009; Strickland *et al.*, 2010), disregard collisions in their explanations (Popova and Bretz, 2018; Sweeder *et al.*, 2019) and/or hold deterministic mental models (Garvin-Doxas and Klymkowsky, 2008; Taber and García-Franco, 2010; Talanquer, 2013; Yan and Talanquer, 2015). As a solution to these challenges, a potential tool that can be used to help students build a scientifically accurate mental model of the sub-microscopic level (i.e., dynamic, collision-based, and probabilistic) is external representations, such as pictures or animations.

Technology advancements – like the development of animations and simulations – have given learners more tools to construct a more scientifically accurate visualization of the submicroscopic domain. Studies have demonstrated that animations can help learners visualize the sub-microscopic level (Stieff and Wilensky, 2003; Velázquez-Marcano *et al.*, 2004; Tasker and Dalton, 2006; Kelly and Jones, 2008; Kelly and Akaygun, 2016; Kelly *et al.*, 2017, 2017, 2021). However, animations usually focus on successful collisions and as a result may lead to deterministic thinking (Tasker and Dalton, 2006; Dangur *et al.*, 2014), or lead students to focus on non-critical features (Tasker and Dalton, 2006; Bussey and Orgill, 2015). Our goal was to identify participants' mental models of the sub-microscopic level and determine which mental model participants use while engaging with a molecular dynamic visualization.

## 4.9 Theoretical frameworks

### 4.9.1 Mental models

Models are used to communicate, predict, and explain sciences. There are two types of models: conceptual models and mental models. Conceptual models are scientifically agreed-on

models that are used by the scientific community to communicate chemical phenomena (Greca and Moreira, 2000). These models have been studied by scientists and have been agreed upon using the evidence obtained through scientific experimentation.

On the other hand, mental models are internalized representations held by a person that constantly change and are abstract (Greca and Moreira, 2000). Mental models have been defined as "*Internalized, organized knowledge structures that are used to solve problems. They are encoded with respect to the spatial, temporal, and causal relationship of a concept"* (Rapp, 2005). A person will build a mental model around aspects of the physical and social world they know (Bower and Morrow, 1990) and they are not necessarily right or wrong, as they are a representation of an individual's experiences (Johnson-Laird, 2010). Therefore, new experiences, new information, or new situations will often result in changes to an individual's mental models.

A person will have multiple mental models of the same concept, built from experiences and interactions with that concept (Greca and Moreira, 2000). When faced with a problem, one or more specific mental models will be used that are relevant to the problem/context at hand in the working mental model (Figure 4.3) (Johnson-Laird, 2010). This working mental model is temporal and linked to the context it is used in where a person will use only the characteristic(s) they consider necessary (Johnson-Laird, 2010; Kelly, 2014).

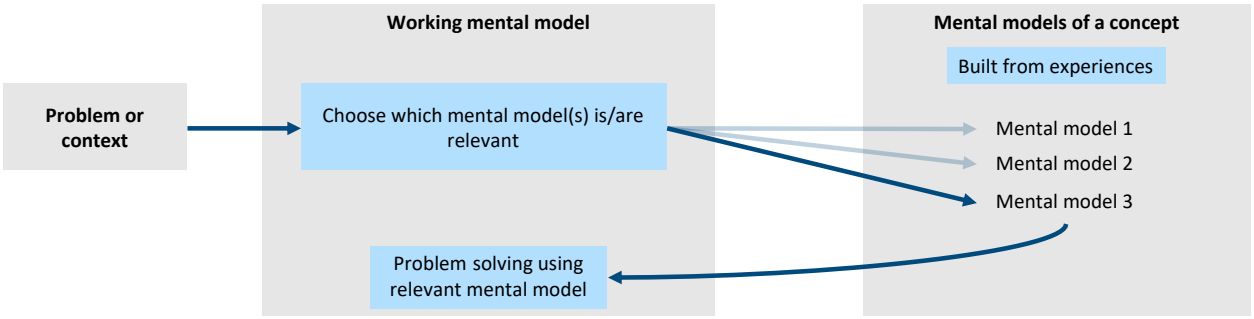



*Figure 4.3. Schematic representation of mental models*

When new information is introduced that does not "fit" with a learner's previous mental model, cognitive dissonance will occur (Festinger, 1957; Cooper, 2019). The learner will typically experience discomfort and try to resolve the variation between the new information and their previous mental models by either rejecting the new information or adapting their existing mental model to include the new information. This can be done by choosing to dismiss the information

as unimportant or attempting to alleviate the dissonance by either changing the learner's opinion or initiating a debate to attempt to change the researcher's opinion (Festinger, 1957; Aimeur *et al.*, 1997). The introduction of cognitive dissonance in learners has led to changes in mental models if the participants choose to use the information (Heemsoth and Heinze, 2014).

## 4.10 Research questions

Since a robust mental model of chemistry includes a dynamic understanding of how molecules behave and interact (Bodner and Domin, 2000), our goal is to identify students' mental model of the sub-microscopic level. Molecular dynamic visualization can provide the tools to help students visualize a dynamic, collision-based, probabilistic mental model (Tasker and Dalton, 2006). Therefore, we can determine which mental models students will use before and after using a molecular simulation. In this study we addressed the following research questions:

**Research question 1:** What mental models are participants using before and after viewing a molecular dynamic visualization?

**Research question 2:** How might a molecular dynamics simulation affect conflicting mental models of randomness?

## 4.11 Method

### 4.11.1 Participants

This research was approved by the institution's Research Ethics Board before any participants were recruited (H-03-20-5585). Four participants (P1–P4) were recruited from a second-year biochemistry course. As the data was analysed, we found that code saturation was not obtained, and more participants were required. A second recruitment round was done from a first-year organic chemistry course where 6 participants (P5–P10) were recruited to obtain code-saturation. One participant (P5) did not finish the intervention and therefore, their data was removed from the study.

#### *4.11.1.1 Study overview*

The study consisted of a 2-hour Zoom semi-structured interview with an experimental design (i.e., data collection before and after the intervention). Participants were invited to participate in a Zoom interview via a posting on their classroom Learning Management

(Brightspace) page. The participants signed the informed consent form before the study commenced (Figure 4.4).

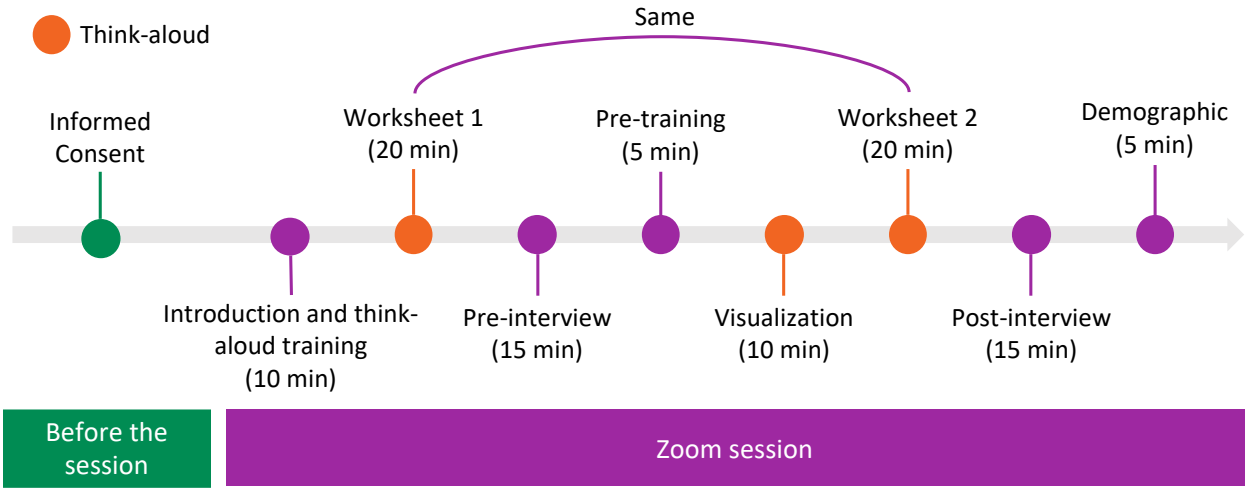



*Figure 4.4. Overview of the study*

Once the Zoom session started, participants were welcomed and trained on the think-aloud protocol (Birch and Whitehead, 2019). A think-aloud protocol (i.e., where participants vocalize their thought processes while they answered the worksheet and while viewing the visualization) was added after the first four participants were interviewed, and therefore this data was only used for the last five participants. Participants were trained by asking them to draw and verbalize something they had done the previous day. Once trained in think-aloud protocols, the participants were asked to fill out a worksheet that contained six questions while verbalizing their thoughts (Appendix 1.1). The participants sent their worksheets electronically to the interviewer and were then interviewed for 10 to 20 minutes (pre-interview phase).

After the pre-interview, participants were given a pre-training video to orient them in the simulation. The participants requested control of the interviewer's screen and could manipulate the simulation as they wished for as long as they wanted (on average 5 min). Once finished with the visualization, the same worksheet was given, and the same interview protocol was followed. At the end of the study, a demographic survey was administered asking about the participants' prior classes and years of study.

#### *4.11.1.2 Odyssey by Wavefunction*

In our study, we used two simulations from the Odyssey program created by Wavefunction (Wavefunction, 2019). The participants took control of the interviewer screen to interact with the molecular dynamic simulation. The students were given a pre-training video outlining the actions they could control in the simulation (e.g., starting the simulation, zooming in and out, rotating the cell, altering the speed of the simulation, changing the representation, and going through the simulation frame-by-frame).

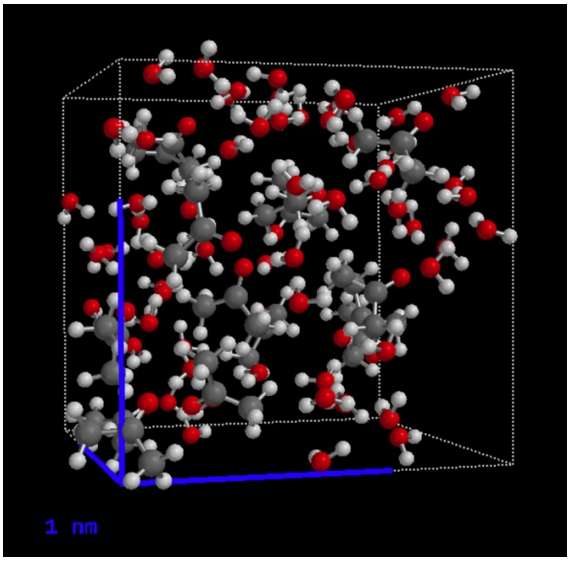
The first simulation was a solution of acetone in water (Figure 4.5a, Left), which was chosen to demonstrate a liquid solution and the relative distance between molecules. The simulation contained no reactions but demonstrated both internal motion (i.e., bond rotation and vibration) and displacement through space.

The second simulation was a reaction between fluorine and $NO_2$ in the gas phase (Figure 4.5b, Right), which was chosen because it shows collisions between molecules at equilibrium. The simulation showed that collisions are completely random and probabilistic. In order to help participants see collisions the simulation highlighted them in grey when the simulation was being played.

a) A solution of acetone in water (no reaction).

b) A gas phase simulation between $NO_2$ and $F_2$.

$$F_2 + NO_2 \leftrightharpoons NO_2F + F$$
$$NO_2 + F \rightarrow NO_2F$$



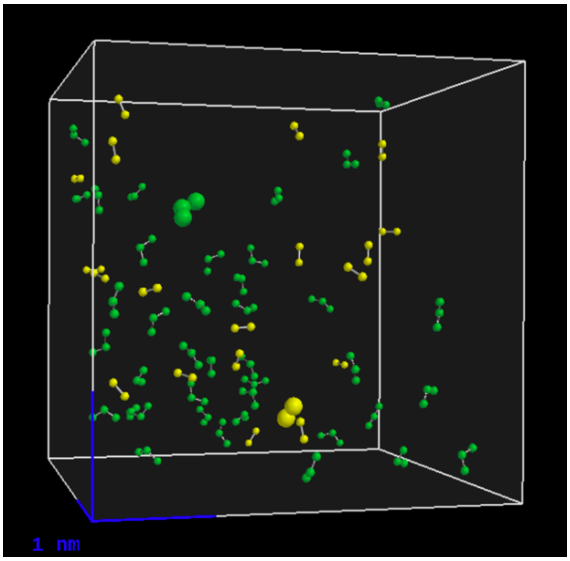




*Figure 4.5. Screenshot of the simulation viewed by the students. Left: Acetone and water mixing, right: Fluorine and $NO_2$ reacting inside the cell.*

### 4.11.2 Data collection

Data came from three sources: the worksheets, the think-aloud interviews, and the semi-structured interviews. These three sources of data were used to triangulate our findings and increase the trustworthiness of our findings (Creswell and Miller, 2000). The worksheet consisted of six questions and can be found in Appendix 1.1. The interviews were semi-structured, which means the interviewer asked the participant prompts and followed up on other specific concepts (Box 4.1). When the interviewer followed up on a comment by the participant, they did not introduce new concepts and as such some concepts may not have been mentioned by all the participants.

**Interview prompts**

- Can you explain your first drawing to me?
- Was there anything you couldn't capture in your drawing?
- What are the molecules doing overtime?
- Can you explain your second drawing to me?
- Was there anything you couldn't capture in your drawing?
- For the last question you circled [1-10], can you explain why?
- Post-interview only: What did you think of the simulation?
- Post-interview only: Was there anything you wish the simulation showed?

**Follow-up questions**

- You mentioned [motion/amount/collision], could you expand on what you mean?
- Could you elaborate on the answer you just gave?
- You just mention [motion/amount/collision], what do you imagine that in your head?

*Box 4.1. Interview guide including the prompt and sample follow-up questions.*

### 4.11.3 Data analysis

The interviews were transcribed verbatim using Otter.ai software and edited for accuracy by the first author. All analyses were conducted in NVivo 11 for Mac.

To address research question 1: "*What mental models are participants using before and after viewing a molecular dynamic visualization?"* we used deductive coding (Gibbs, 2018) using three types of mental models previously outlined by Bongers et al. (2019) as the starting codes: static, process, and particles in motion. *Static* mental models were coded as participants describing their drawings. *Process* mental models were coded as the participants describing a transformation as a process, without mentioning a transition or change. *Particles in motion* was coded as participants visualizing molecules as dynamic, including transition. Following the first round of deductive coding, the authors expanded the codebook to incorporate codes related to the three general concepts of motion, amount, and collisions. The codes generated through the deductive coding were then used to create categories and themes around participants' dynamic mental models.

To address research question 2 "*How might a molecular dynamic simulation affect conflicting mental models of randomness?*", we used inductive open coding, using a constant comparative analysis method, i.e., re-coding previous interviews with new codes (Corbin and

Strauss, 2008). This round of coding was done using gerund verbs to describe the actions and beliefs of the participants (Charmaz and Henwood, 2017), through the lens of the research question (e.g., “Visualizing molecules as floating”, “Using polarity to explain why molecules move”). Through constant comparative analysis, categories were developed from the codes (Corbin and Strauss, 2008; Saldaña, 2013). We then focused our analysis by asking questions of the data and narrowing the focus of the codes and categories (Ryan and Bernard, 2003). The following questions were asked to the data: What are participants’ mental models of randomness? How are the participants using the simulation to learn?

#### *4.11.3.1 Reliability*

To establish reliability a second coder independently coded 15% of the data. The codes were compared; 87% percent agreement was reached and the Krippendorff a was 0.89. Both of these are considered above acceptable limits, which are > 0.7 for a (Krippendorff, 2004) and > 80% for percent agreement (McHugh, 2012).

## 4.12 Results and discussion

### 4.12.1 Students have a dynamic mental model of the sub-microscopic level

All participants used the three mental models outlined by Bongers et.al. (2019); with the particles in motion being the most frequent mental model (Figure 4.6). During the post-interview, participants used a *particles in motion* mental model more often than in the pre-interview. The *process* mental models had the opposite trend (i.e., less in the post-interview). There was little to no change in the *static* mental models. Participants may have used a *particles in motion* mental model because the simulation may have cued them. There may also be some participant bias, in which the participants are trying to “guess” what the interviewer wants to hear, in the post-interview (Gove and Geerken, 1977; McCambridge *et al.*, 2012). While the interviewer tried to be neutral and not cue the participants, the simulation itself may have cued a *particles in motion* mental model. However, previous research showed that participants used the animation as an inspiration for their answers (Kelly, 2014) and that molecular dynamic visualization does promote dynamic mental models (Williamson and Abraham, 1995; Akaygun, 2016; Kelly *et al.*, 2017).

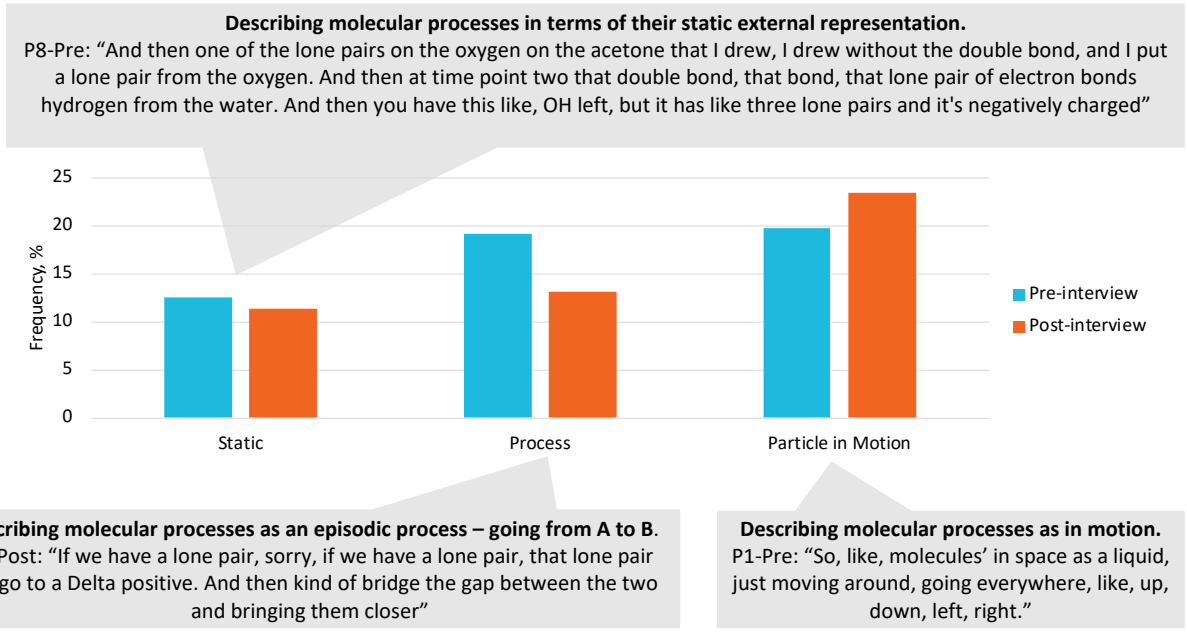



*Figure 4.6. Frequency of static, process, and particles in motion mental models as a percentage of all mental models used (N = 167)*

Process mental model has been reported in students' reasoning about mechanism and reactivity (Galloway *et al.*, 2017, 2019; Bongers *et al.*, 2019; Graulich and Caspari, 2021), and it was also present in the current study despite one of the simulation not showing a process (i.e., reaction) occurring. Participants used the *process* mental model more frequently when describing a reaction (i.e., the second simulation). Also, the participants explained motion as a process by talking about how molecules change position without specifically saying they are moving. Participants using a *process* mental model for a non-reaction accounted for 4% of all mental models seen in this study.

The *particles in motion* mental model included participants who talked about the transition between two points (Bongers *et al.*, 2019). This mental model has been reported as absent when students discuss reaction coordinate diagrams (Popova and Bretz, 2018; Atkinson *et al.*, 2021) or mechanistic problems (Bhattacharyya and Bodner, 2005; Ferguson and Bodner, 2008; Strickland *et al.*, 2010). However, we found that within the context of using a molecular dynamic visualization, participants used this mental model readily. Within this mental model there were two components: motion and collision (Figure 4.7). Each contained sub-categories of

mental models, and participants used a variation of those mental models. Participants also demonstrated different mental models about the number of molecules that were present.

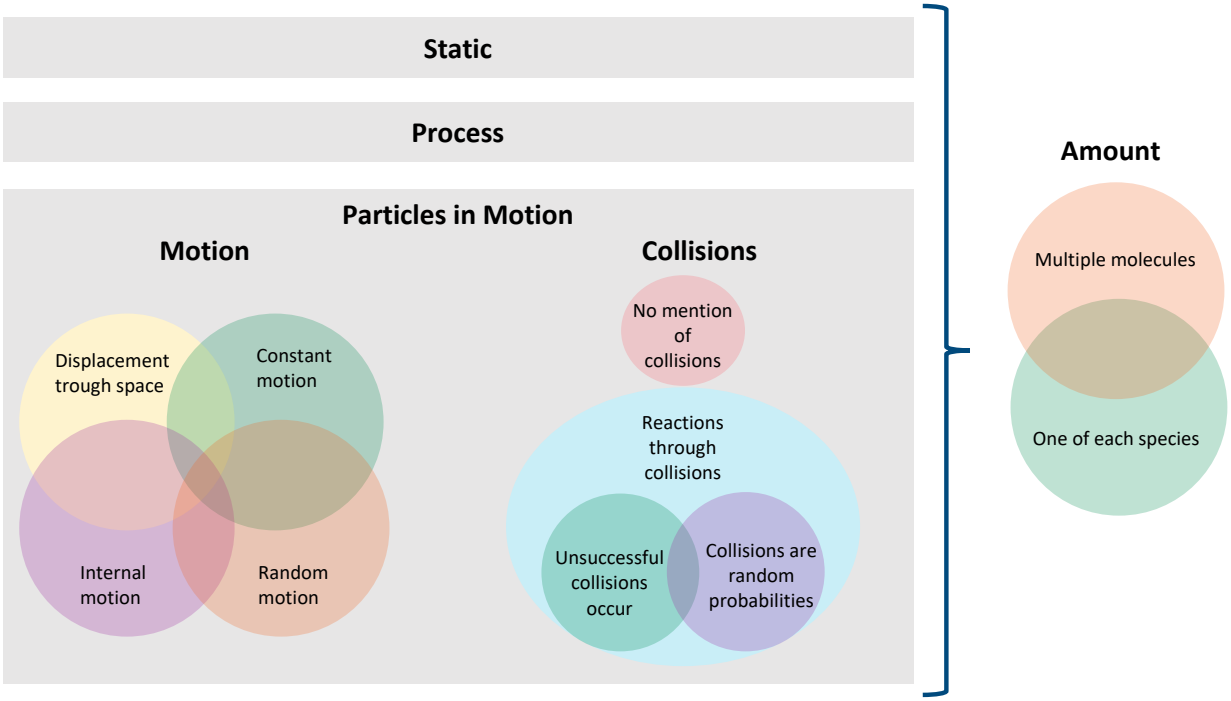



*Figure 4.7. Overview of the mental models of the sub-microscopic level. The size of the circles does not have meaning.*

#### *4.12.1.1 All participants mentioned displacement through space*

The motion mental model was divided into four sub-mental models: (1) *displacement through space*, (2) *internal motion*, (3) *constant motion*, and (4) *random motion* (Figure 4.8). The four mental models could overlap and were not mutually exclusive.

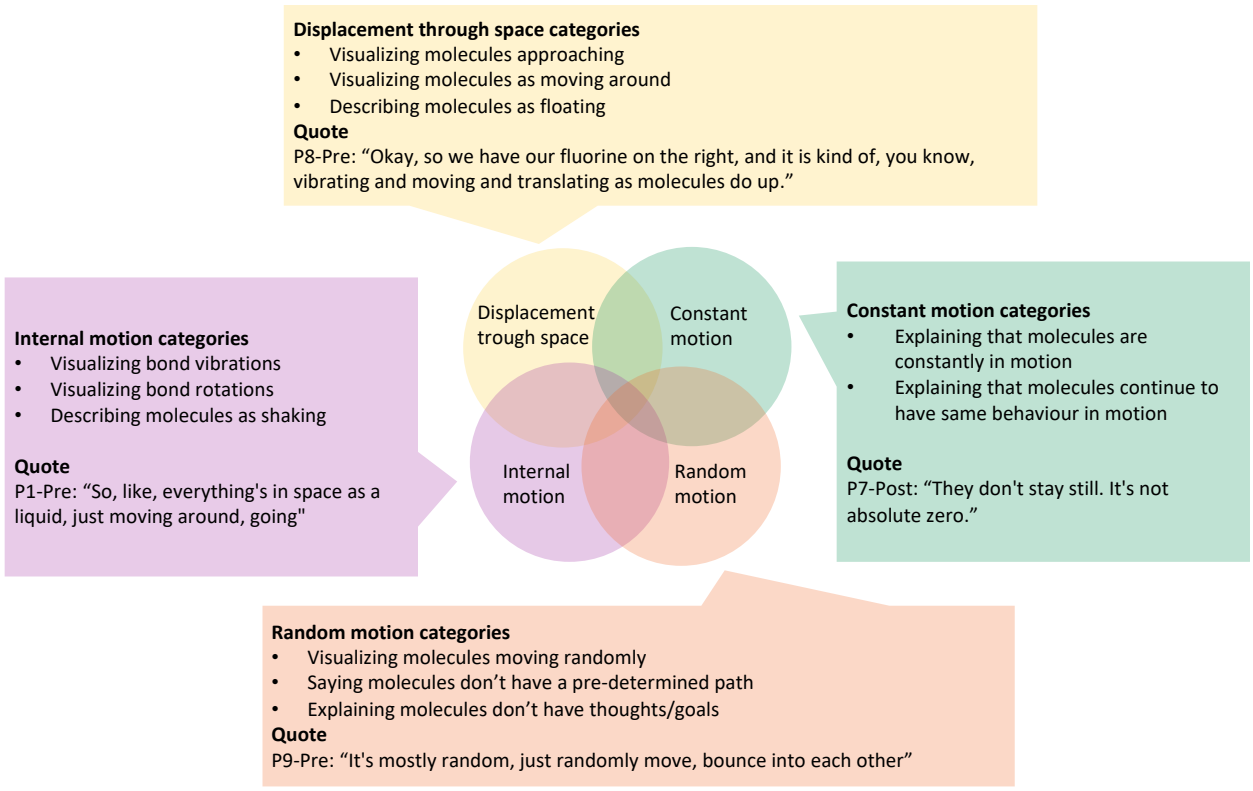



*Figure 4.8. Mental models of motion with associated codes and representative quotes*

All the participants (both pre- and post-interview) displayed the *displacement through space* mental model by explaining that molecules were translating. Motion was either mentioned outright (Pre: 8, Post: 9) or off-hand phrases such as "free-floating fluoride" or "the molecules approach each other" were used by participants (Pre: 3, Post: 5). Previous work has shown that participants often use a *static* mental model of the sub-microscopic world (Akaygun, 2016; Bongers *et al.*, 2020) but that students gain a more *dynamic* mental model after viewing a visualization (Tasker and Dalton, 2006, 2008; Sanger *et al.*, 2007; Tang and Abraham, 2016). Some of the participants in these studies were in secondary school, which may account for the higher learning gains; however, our results show that university students do have a dynamic, motion-based mental model of the sub-microscopic level.

The *internal motion* mental model refers to bond rotations and vibrations of molecules. Three participants showed this mental model (Pre: 3, Post: 2). P8 and P10 demonstrated the mental model in both pre-interview and post-interview while P3 used it only in the pre-interview. Our results are similar to previous studies where participants focused on general features (i.e.,

translation) as opposed to more detailed features (i.e., vibration and bond rotation), with participants sometimes dismissing them as unimportant (Kelly, 2014). The translation through space is typically more visible in visualizations, and the internal motion could be overlooked. In the think-aloud interview, P8 mentioned that they liked the zoom in/out feature of the simulation, because it allowed them to see the bond vibration and rotation. Without that feature, this participant would not have seen this property. We investigated what participants did while participants viewed the visualization and observed that most participants zoomed in and out, but only P8 zoomed until an entire molecule filled the screen to observe the internal movement.

The *constant motion* mental model is used when participants mention that motion is constant and never stops. Three participants showed this mental model (Pre: 2, Post: 1). Like the *internal motion* mental model, all participants who displayed this mental model also displayed the *displacement through space* mental model and P8 showed all three.

*Internal motion* and *constant motion* mental models were not the focus of the tasks, and therefore the participants may have the mental model but did not use it in this context. For both mental models, the participants brought up each mental model once (except P8 who brought up internal motion twice) during the entire interview, while *displacement through space* was used on average eight times per interview.

The final category of motion mental models is that *motion is random*. Five participants (Pre: 3, Post: 4) explicitly stated that the motion of molecules was random. The participants would explain that motion does not have a pre-determined path, or that it is randomly moving. Others demonstrated the opposite mental model and used teleological (i.e., atoms have wants and goals) or that molecules had a purpose, similar to previous work (Talanquer, 2007, 2013). This mental model will be expended on in future sections.

Overall, all participants held the mental model that molecules moved around through space. A few mentioned that internal motion was occurring, that motion was constant and/or that motion was random. The focus was on the translation of molecule, which is unsurprising since it is the type of motion that is most visible in the simulation.

#### *4.12.1.2 Most participants used collisions to explain a reaction*

Participants displayed one or more of the following mental models (Figure 4.9): (1) *no mention of collisions*, (2) *reactions occur through collisions*, (3) *unsuccessful collisions occur*, or (4) *collisions are random probabilities*. Eight participants used a collision mental model at least once, in the pre-interview or post-interview.

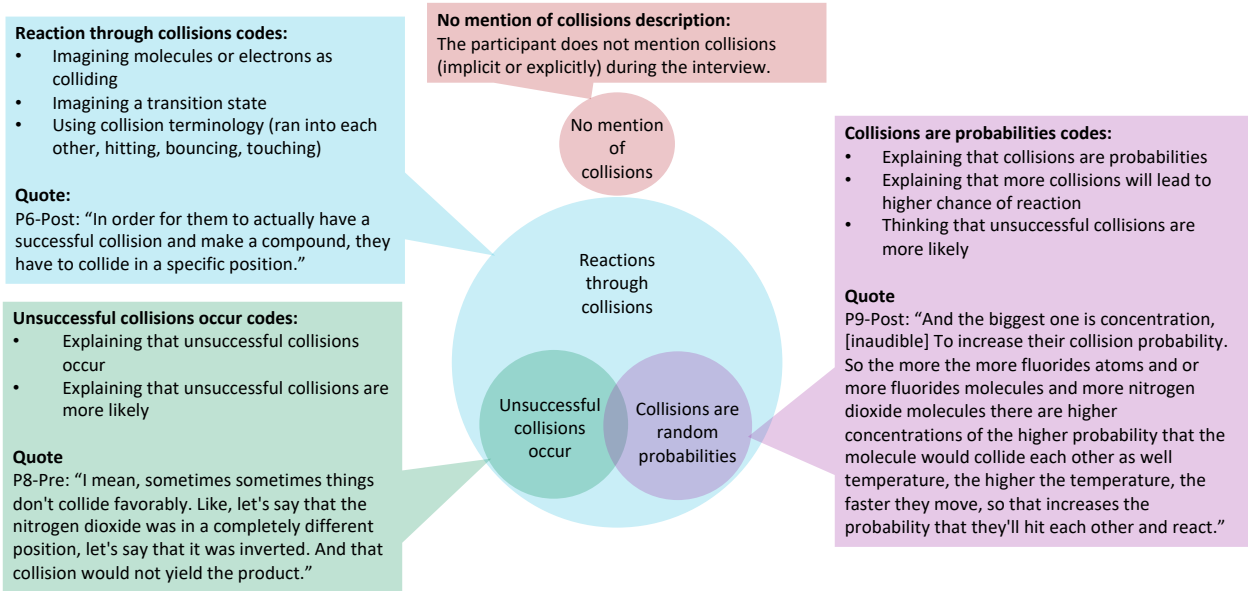



*Figure 4.9. Mental models of collisions and associated child codes and representative quotes*

The first mental model related to collisions is talking about reactions without mentioning that collisions must occur. The participants in this group (Pre: 3, Post 3) mentioned the process of a reaction in terms of electron movement or bonds breaking/forming but did not mention collisions occurring. Some of those participants (Pre: 2, Post: 2) used the word "interaction" which could either mean a reaction or other concepts, and it is unclear if they meant collisions specifically. Participants may have felt they did not need to mention collisions in their explanation or repeat their previous explanations from the pre-interview/think-aloud. While they may not have explicitly talked about collisions, it does not mean that the participant did not have a collision-based mental model. Instead, participants may have thought that collisions were not relevant at that time.

The second mental model is *reaction through collisions* (Pre: 5, Post: 5), meaning that the participants explained that a collision is necessary for a reaction to occur. Previous studies that

investigated students' skills using reaction coordinate diagrams showed that students do not mention collisions often (Popova and Bretz, 2018; Atkinson *et al.*, 2021). However, other studies about energy and pressure saw the majority of participants mention collisions (Macrie-Shuck and Talanquer, 2020). Therefore, the context matters, and in our study most participants explained that collisions do occur for a reaction.

Embedded within the *reaction through collisions* mental model were two mental models related to probabilistic thinking: talking about *unsuccessful collisions* and thinking of *collisions as random probabilities*. These mental models can be difficult to teach, especially since animations have been used to show collisions but tend to only focus on successful collisions (Tasker and Dalton, 2006, 2008). Six participants (Pre: 4, Post: 3) explained that unsuccessful collisions occur in the context of meeting the requirements for a reaction (i.e., orientation and energy). In previous studies, students showed higher skills on a collisions theory and were better at explaining that orientation and energy are required for a reaction to occur (Sweeder *et al.*, 2019). However, in this visualization study, the visualization had no effect on participants' explanation. When probed by the interviewer, three participants (Pre: 3) mentioned that unsuccessful collisions were more likely than a successful collision. Other participants stated they were unsure, or that there would be roughly equal collisions.

The other probabilistic collisions mental model is that *collisions are probabilistic* and that *collisions are random*. Three participants clearly stated that reactions were probabilities of successful collisions. P4's explanation was unexpected since they used probabilistic terminology but did not mention collisions in their explanations. This participant may have felt it unnecessary to mention collisions in their answer while explaining that reactions were a probability. In the quote below, P4 explained that for a reaction to occur the molecules must be in a specific orientation and that adding more molecules would alter the probability of a reaction, without mentioning collisions.

> *P4-Post: Yeah, so the question was asking about molecular requirements [for a reaction]. And I think the only requirement is just the orientation of the reactants, the orientation of the reactants when they*

*interact. Yeah, matter of probability when it comes to molecularity. The increased molecularity makes the reaction less probable, it doesn't make it impossible. So if the orientation is there, anything can happen.*

#### *4.12.1.3 Most participants used different collision mental models before and after the visualization*

Five participants changed the mental model they used from the pre-interview to the post-interview (Figure 4.10). None of the mental models were more prevalent in the pre- than in the post-interview (or vice versa). Possible reasons why participants used different mental models could be that they did not want to repeat themselves, and/or they felt the mental model was not relevant to the context, and/or they did not hold the mental model. While we do not know the specific reason(s) for mental models used by each participant, we can conclude that participants use and demonstrate collisions-based mental models.

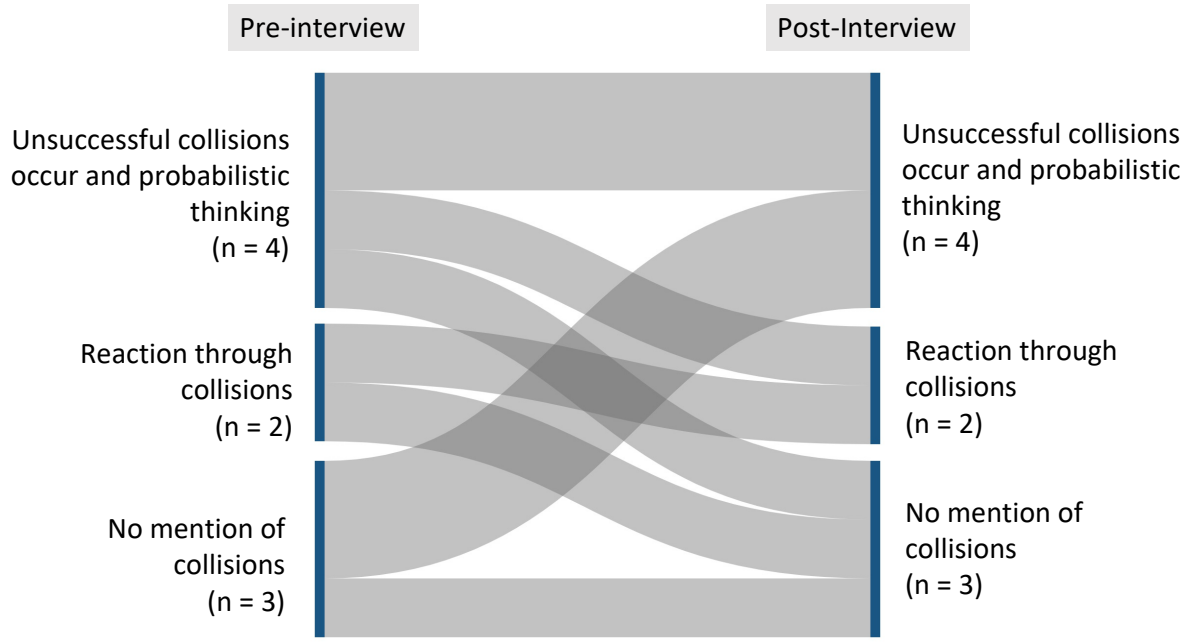



*Figure 4.10. Mental models that participants displayed related to collisions from the pre-interview to the post-interview.*

#### *4.12.1.4 Participants often focused on one of each species, but also demonstrated they knew many molecules were present*

Each mental model outlined (*static, process,* or *particles in motion*) could be uses in conjunction with two mental models: (1) *one molecule of each species*, or (2) *multiple molecules*.

The *multiple molecules* mental model (i.e., visualizing many molecules of multiple species) was demonstrated by all participants (in both pre-interview and post-interview). This differs from previous literature where students tend to focus on the reacting species, and ignore solvents or molecules not involved in a reaction (Popova and Bretz, 2018). This could be related to the prompt itself; Popova and Bretz (2018) investigated students' skills using a reaction coordinate diagram, while the prompt for our work was a storyboard and focused more on participants' mental model of the sub-microscopic level. Other studies that used storyboards, showed that participants used multiple molecules in their drawings (Tasker and Dalton, 2006, 2008; Akaygun, 2016; Kelly *et al.*, 2017). Similarly, students in previous studies stated they understood why an animation would remove water (i.e., for simplicity) (Kelly and Hansen, 2017).

Comparing the amounts of molecules mental models (i.e., *one of each species* vs. *multiple molecules*) to the *static, process,* and *particles in motion* mental models showed that the *multiple molecules* model is most often used with a *particles in motion* model. While *process* and *static* mental models were used most often with the *one of each species* mental model (Figure 4.11a). Most participants (Pre: 8, Post: 6) drew a single molecule of each species on their storyboard and then used the one of each species mental model for their explanation, which was typically a *static* mental model.

Participants shifted to the *multiple molecules* mental model in the post-test (Figure 4.11b). While all participants mentioned *multiple molecules* in the pre-interview, the simulation could have led them to use the mental model more often in the post-interview.

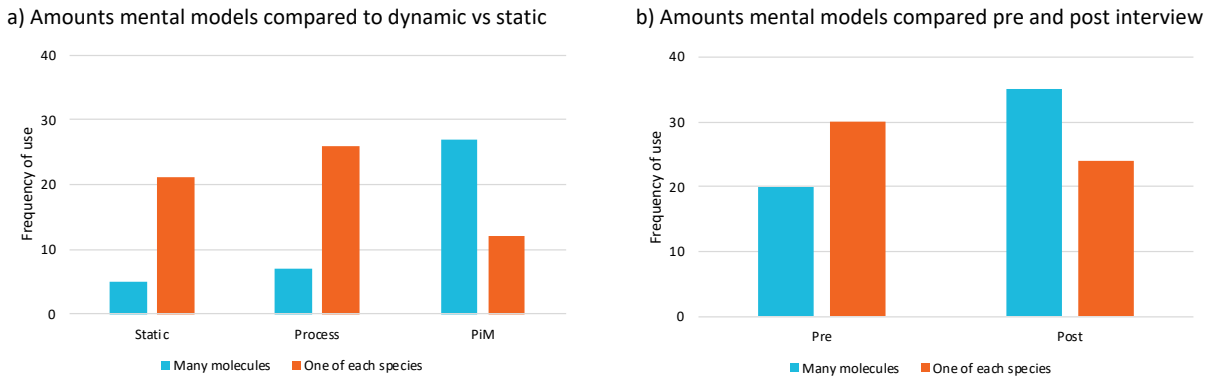
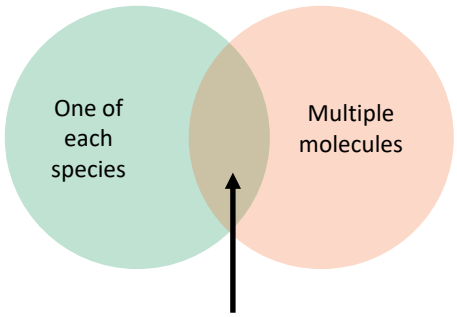



*Figure 4.11. Mental models related to the amount of particles present compared to (a) static vs dynamic and (b) pre vs post.*

Some participants (Pre: 1, Post: 5), stated they knew there were many molecules in the solution but chose to only visualize two (Figure 4.12). Some participants stated it was for simplicity while others stated they did not feel like drawing all the molecules. For example, multiple participants (P1, P3, P7, P8, P10) stated that they were aware of the multiple molecules present but chose to focus on one of each species to visualize an interaction. Our work shows that participants have multiple mental models of a solution (one vs multiple) and choose one of the mental models over another for the task at hand.




P8-Pre: I know that there are like plenty of other molecules in the solution, like there's loads, and they're all sort of doing the same thing. But when I'm picturing a single interaction, I like looking at it between the two molecules, because I know that there would be lots of other molecules surrounding them."

*Figure 4.12. Participants at the junction of two mental models*

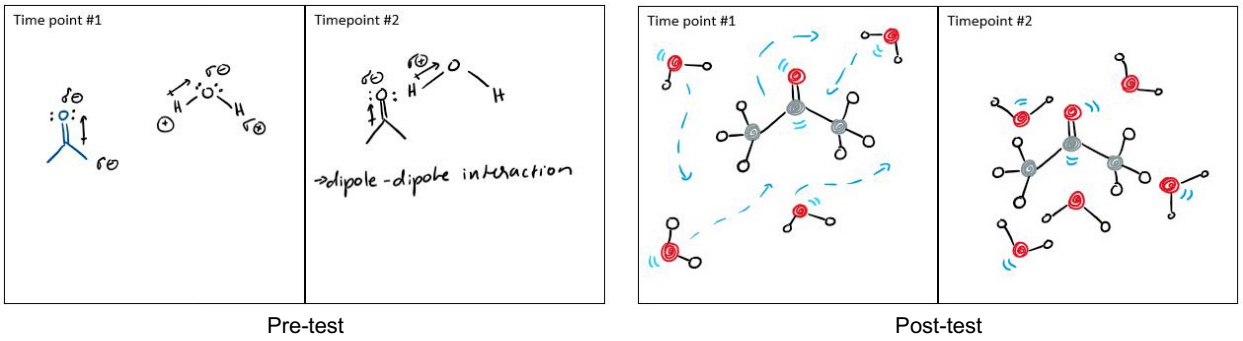
### 4.12.2 Participants used different mental models depending on the context

The *static* mental model was used by all participants when explaining their drawing and amounted to 60% of all static mental models used when answering the question "Can you explain your drawing to me?". Through discussion (i.e., during the semi-structured component of the interview), participants used a particles in motion mental model 57% of the time. The prompts "What are molecules doing over time?" and "Can you explain why you put [1-10] on the scale of random to purposeful?" led to participants using mostly a particles in motion or process mental model. Bongers, Northoff, and Flynn (2019) had similar findings, indicating that the mental model participants used was prompt dependant, and our current prompt may have prompted the students to simply explain their drawing.

For the "Was there anything that you couldn't draw, or was challenging to draw?" prompt, three participants (Pre: 2, Post: 2) said that there wasn't anything missing, or that it was

not challenging. Others (Pre: 2, Post: 2) included answers that were related to their drawing abilities, as opposed to what they thought. Four participants mentioned that they struggled to show what they were thinking but that the simulation showed it better.

Participants (Pre: 4, Post: 3) had discrepancies between their drawings and explanations of their visualizations. For example, P8-Pre drew a Lewis structure (Figure 4.13) in their storyboard but explained in the interview that they visualize molecules and interactions in terms of orbitals. P8 also used a motion-based mental model in both the pre-interview and post-interview but they drew a static Lewis structure with dipole moments in the pre-test (Figure 4.13). In the post-test, they drew a drawing that demonstrated motion (blue lines).



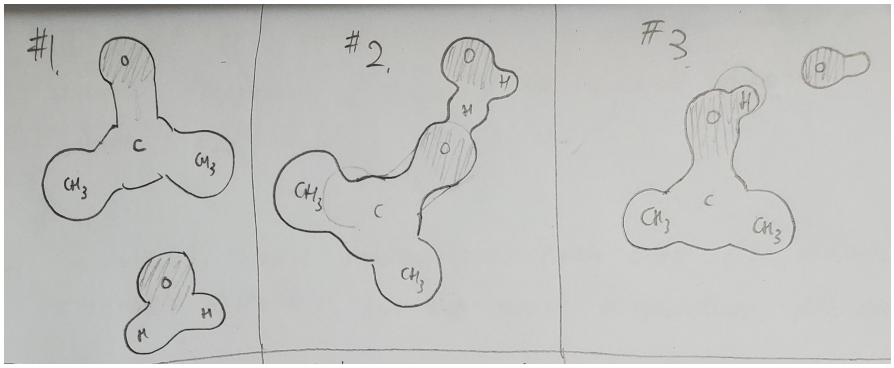

P8-Pre: "So when I think of compounds interacting, I just think of overlapping orbitals sort of hold the entire thing together. So that's what I would imagine an interaction between the two of them to be."
I: "So you just mentioned that you would think of them more like orbitals? Is that what like your first visualization goes to?"
P8-Pre: "Yes. Yeah. Yeah, just a bunch of electrons, like moving and then overlapping with one another."

*Figure 4.13. Example from P8 of a change in drawing, with the associated quote that they visualize the sub-microscopic level differently than drawn.*

Some of the participants (Pre: 3, Post: 3) explained that the situation would lead them to use different representations and mental models. The effect of context on students' sub-microscopic explanation has been previously investigated and showed that students would use different and sometimes contradicting explanations depending on the context they were studying (Teichert *et al.*, 2008). In our study, participants reported that the use of the Lewis structures was prompted by the visual representation and the "school-like" setting.

All the participants' drawings used the Lewis structure representations in their pre-test and four participants mentioned it was due to familiarity and two participants mentioned the questions prompted the use of that representation. The familiarity of a representation has been investigated in previous research and found that the Lewis structure was typically preferred by learners even if a different representation was more useful for the problem (DeCocq and Bhattacharyya, 2019). The question in this study included a Lewis structure of acetone and may have led to participants being cued to use it. However, when asked to draw about dissolution students used balls in their storyboard (Kelly, 2014) and similarly when asked about phase change students drew more particulate drawings (Ardac and Akaygun, 2004). The previous study used dissolution or phase changes as opposed to an organic reaction.

For the post-test three participants chose to use a different representation. Two of those participants who changed their representation changed it to the ball and stick model. P6 mentioned that the simulation had prompted them to change their drawing. P3 drew their post-test storyboard with what they called "blobs" (Figure 4.14). The participant mentioned during their pre-interview that they couldn't deviate from the Lewis structure representation but that they visualized molecules more as electrostatic maps.




P3- Pre: "I think it depends on the context. If it's anything school related, then I would probably think more in terms of like the Lewis sticky drawings with the with the arrows and everything, just because like I have to think that way for school, but you know, if I was in a conversation or just in some social setting, um, you You know that I think that would really change. And it would become more of those, those pictures that you see in the textbooks where, you know, like, the red part is where the electrons are, and then the blue parts where there's less of them, and then they're like, they're these bubbly figures. So that's, that's what I would picture if it weren't for school"

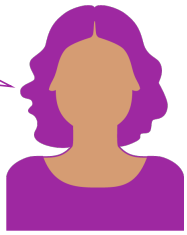


*Figure 4.14. Post-test drawing of P3 using a unique representation, with an associated quote.*

### 4.12.3 Probabilistic and deterministic mental models

Probabilistic thinking in chemistry is thinking that reactions and motion in the sub-microscopic level are random and that collisions are probabilities, which can be difficult to understand for learners (Garvin-Doxas and Klymkowsky, 2008; Jolfaee *et al.*, 2014; Savard, 2014) (Figure 4.15). Learners using deterministic thinking tend to (1) use attraction or external forces to explain how a reaction begins (Yan and Talanquer, 2015), (2) explain that a reaction requires a cause to occur (Taber and García-Franco, 2010; Yan and Talanquer, 2015), (3) use teleological arguments (i.e., molecules want or have goals) (Talanquer, 2007, 2013; Yan and Talanquer, 2015), or (4) explain a reaction will occur because there is an active (i.e., reactive) species (Taber and García-Franco, 2010; Yan and Talanquer, 2015). However, the conceptually accepted mental model is probabilistic where only probabilities and randomness affect the sub-microscopic level.

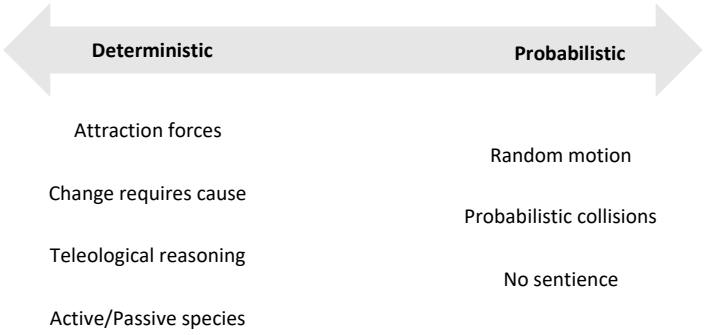



*Figure 4.15. Deterministic and probabilistic mental models*

From our analysis, three categories of mental models emerged as demonstrating probabilistic thinking: (1) *random motion* (2) *collisions probabilities*, and (3) *no sentience*. The first two mental models were discussed earlier, and the third mental model is that molecules are not sentient (Pre: 3, Post: 2). These participants explained in various ways that molecules do not have wants or desires and that molecules cannot think (Taber and García-Franco, 2010; Yan and Talanquer, 2015).

The language educators use when talking about molecules and the sub-microscopic level can often be teleological (Talanquer, 2007, 2013). Teleological explanations are focused on the goal/purpose of something, as opposed to focusing on the cause of an event (Wright, 2020). In chemistry, teleological explanations are present when adding purpose to a chemical system or

explaining why particles react in a specific manner or giving personification attributes to molecules (Figure 4.16)(Talanquer, 2007).

Four participants referred to molecules as wanting something, being comfortable, or being happy. When the interviewer questioned P3-Post about describing the molecule as "being happy", they explained that it was a synonym to the word "stable". When questioned about the use of personification words, P8 simply stated that the personification words they used did not represent wants and that it was a bad semantic. The explanation of using a bad semantic was seen in three of the four participants who used personification. In this case, the participant was aware that they used personification and that the personifications were not accurate. However, three participants believed that molecules have a purpose to react, but none could explain where the purpose came from and that it was just there.

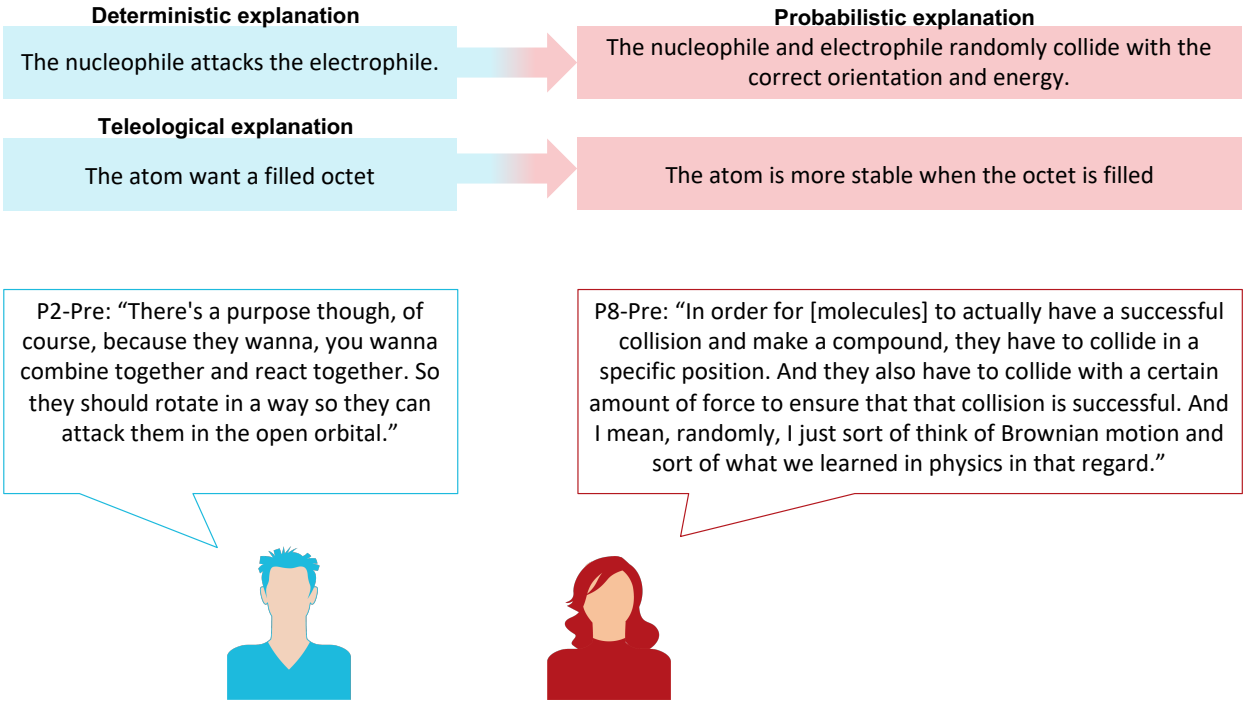



*Figure 4.16. Example and participant quotes of deterministic and teleological explanations (left, blue) and of probabilistic explanations (right, red).*

#### *4.12.3.1 Participants have conflicting mental models about randomness and probabilities*

Only one participant believed the sub-microscopic level was completely random, while only one participant believed it was purposeful. The participant that showed a purposeful mental

model in the pre-interview incorporated more random beliefs following the simulation. The idea of extremes seems to also promote conflicting mental models. P8 was struggling to come to terms with the "extremes" of the randomness scale stating they could not accept that either extreme was correct. Most participants (Pre: 7, Post: 8) showed a mixed perception of the randomness of the sub-microscopic level; they believe there is both purpose and randomness. Previous work showed students struggling to understand and conceptualize randomness in biology (Doerr, 2000; Garvin-Doxas and Klymkowsky, 2008), and in chemistry (Taber and García-Franco, 2010; Yan and Talanquer, 2015).

All participants mentioned randomness but then used conflicting mental models later in the interview. This is similar to previous studies in biology where over 95% of responses said the diffusions process was random but did not display random thinking (Garvin-Doxas and Klymkowsky, 2008). Four conflicting mental models of randomness were used by the participants: (1) adding a driver, (2) rationalizing or being unable to explain randomness, (3) explaining that because the reaction occurs and are an efficient process it can't be random, and (4) that molecules have a purpose (Figure 4.17).

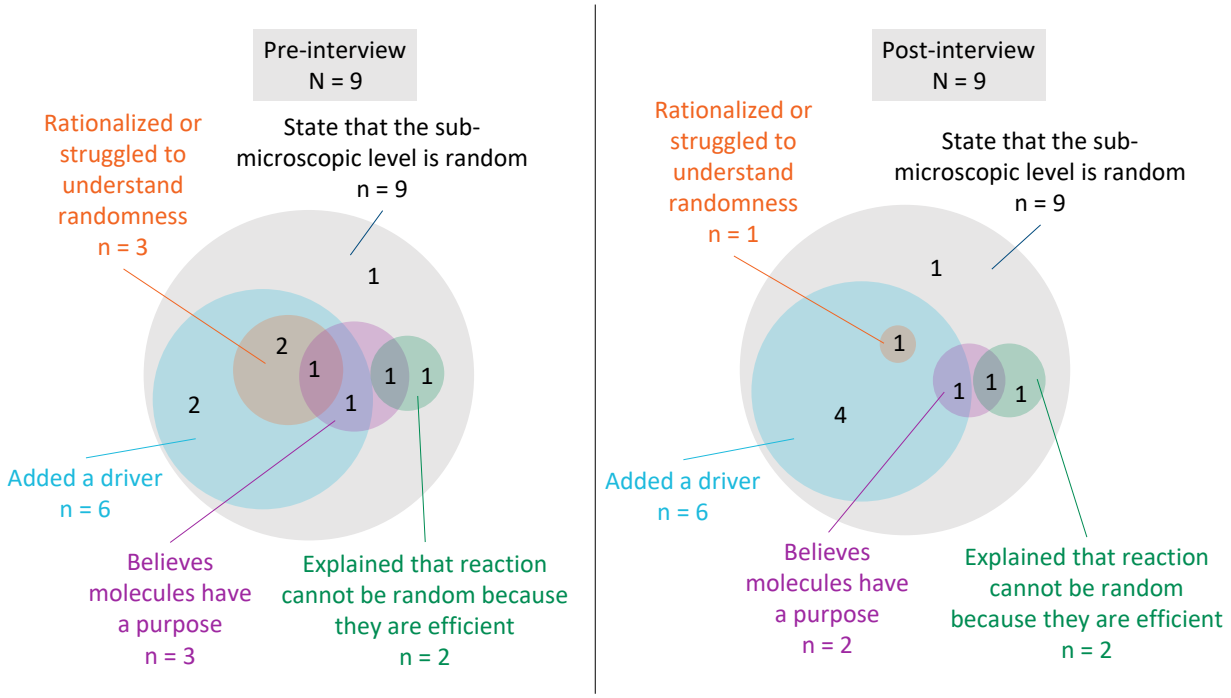



*Figure 4.17. Mental models of participants about randomness*

Many participants added drivers (or cause) to the process, meaning that some properties or conditions can make the molecular world less random. The addition of a driver can be a type of resolution in cognitive dissonance because they add the "*but only"* (Cooper, 2019). The *but only* statements occur when a participant was shown the correct answer and was not given the freedom to choose or build the knowledge (Linder *et al.*, 1967), or the dissonance led to unwanted results (Cooper and Worchel, 1970). For example, P4 expressed the belief of "it is random *but only* for non-polar molecules".

The drivers used by students span the macroscopic and microscopic levels. The microscopic drivers include properties of molecules (i.e., charge, polarity), forces (i.e., hydrogen bonds, dipole-dipole forces), proximity to other molecules, and stability. The most used driver was charges (Pre: 8, Post: 7) and was used in the context of affecting random motion due to attraction and repulsions forces. Similarly, forces (hydrogen bonds, dipole bonds, Van der Waals) were used by many participants (Pre: 8, Post: 4). Participants also used macroscopic changes as a cause for a reaction to occur, similar to a previous report where students explained that there must be a cause for a reaction (Taber and García-Franco, 2010). Two participants (Pre only) mentioned that containing the solution (i.e., in a closed container vs. in the air around us) would reduce the randomness and several participants (Pre: 3, Post: 2) explained that chemists can change the randomness of the sub-microscopic level, or that in a laboratory setting the randomness is different. Examples include the idea that chemists have the ability to change the randomness (P7-Post) or that the wind will affect the random motion of particles (P3-Post). In these cases, the participants used macroscopic changes as the cause for a reaction. This is also related to attributing macroscopic properties to molecules, similar to previous studies (Chandrasegaran *et al.*, 2007; Kelly and Jones, 2008; Rosenthal and Sanger, 2012) where participants would give sub-microscopic entities macroscopic properties (i.e., color).

Another conflicting mental model was struggling to understand randomness (Pre: 3, Post: 1). The participants were unsure of the meaning of "randomness" and/or "purpose" and therefore came up with their own definitions. For example, the participants could not explain what it meant to be random. P6-Pre said they could not conceptualize or explain why there would be randomness, and so they decided to simply say it was purposeful. P10-Pre had mentioned that

the sub-microscopic level was random, but their worksheet indicated a 6 on the random to purpose scale. However, when the interviewer then asked why they had circled a six and where the purpose was coming from the participant was flustered and could not explain their thought process.

Finally, some participants used the argument that if a reaction could occur and could be predicted, it could not be random. This argument is similar to a previously reported definition of randomness that if it can be predicted it is not random (Jolfaee *et al.*, 2014) or that randomness is completely inefficient while chemical reactions seem efficient (Talanquer, 2007; Wright, 2020). Our results are similar to previous reports that students failed to explain that specific patterns and behaviours could be an emergent property from random processes (Luisi, 2002; Garvin-Doxas and Klymkowsky, 2008; Talanquer, 2008, 2015). The lack of thinking about randomess could be related to the teaching of the subjects that tend to focus on efficient and successful processes (Tasker and Dalton, 2006, 2008; Garvin-Doxas and Klymkowsky, 2008).

#### *4.12.3.2 Cognitive dissonance from the simulation*

The simulation showed the randomness of the sub-microscopic level, which may have conflicted with learners' prior mental models. The participants displayed three types of cognitive dissonance: (1) interpreting the simulation to "fit" their mental models, (2) dismissing the simulation, or (3) using the simulation to change their initial mental models.

From the beginning, P1 visualized a scientifically accurate model of randomness and collisions, stating that the molecular world is just physics. Upon viewing the animation, they stated they had seen many similar videos in their schooling and were not surprised by the simulation. The participant did not change anything in their worksheets and the post-interview did not show any difference in mental models, as such they are the only participant that did not experience cognitive dissonance.

Participant 2, 4, and 9 used their interpretation of the simulation to fit within their previously held mental models. Their individual interpretation could be a way to alleviate cognitive dissonance, or these participants simply did not focus on the salient features of the simulation. In our case, participants 2 and 4 misinterpreted the motion of molecules within the

simulations. P2 mentioned the motion of molecules demonstrated in the simulation showed that they had a purpose. P2 further stated that the molecules were moving toward one another with the purpose to react. P4 explained that the motion of the molecule was greatly unaffected by the non-polar region of acetone, but that the polar oxygen affected motion. This phenomenon was not demonstrated in the simulation. Both participants' interpretation of the simulation "fit" into their pre-interview mental models.

Participant 9 interpreted the visualization correctly and mentioned random motion, but they believed that the simulation was missing some aspects. The participant held the mental model that molecules sped up when they got closer to one another (which was not seen in the simulation). When asked about what they thought of the simulation, they mentioned they wished that the speeding up process was demonstrated. In this case, P9 believed that the simulation was missing something they previously believed.

Two participants dismissed the simulation to retain their pre-interview mental models. These participants held the same randomness mental model both before and after viewing the visualization. P3 and P4 stated that how they visualized the sub-microscopic level had not changed, which was consistent with their responses in the post-interview showing the same mental model. Both held a random mental model with conflicting mental models of adding a driver, which they demonstrated in both pre- and post-interviews.

Some participants (P6, P8, P10) used the simulation to change their mental models. The biggest change in mental models came from P6 who held the alternate conception that molecules behaved like the Bohr model (i.e., orbiting around a center) in the pre-interview (Figure 4.18). This mental model was already conflicting with other mental models, and the participants could not explain why or how this behaviour occurs. However, after viewing the simulation P6 displayed a random motion and random collision mental model, but still held a non-random mental model related to drivers.

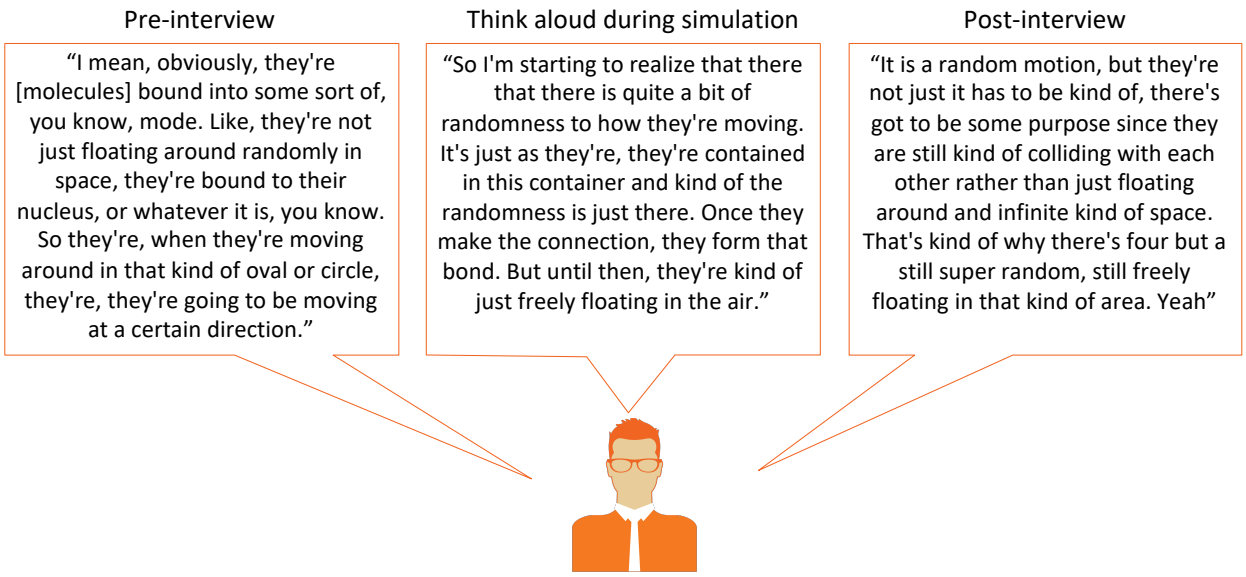



*Figure 4.18. P6's mental models before (left), using the simulation to learn (middle) and a more random mental model (right).*

## 4.13 Conclusions

Our goal was to determine whether the simulation affected mental models used by participants and identify the different types of mental models the participants used. We used an experimental design intervention and interviewed nine undergraduate students, using the Odyssey (Wavefunction, 2019) simulation as the intervention.

Participants have multiple mental models of the sub-microscopic world and use the appropriate mental model for a problem or context. Participants in this study demonstrated both *one of each species* and *multiple molecules* mental models, and then explained that they chose to omit molecules to focus on interactions. Similarly, participants have both static and dynamic mental models but used different ones depending on the context or question asked. All participants used a dynamic mental model that included molecules' displacement through space, and eight participants held the mental model that reaction occurred through collisions at one point during the intervention. Participants also used different mental models about collisions from the pre-interview to the post-interview, either by choice, desire to not repeat their previous explanation, or because the mental model wasn't relevant to them at that point.

Only one participant had a fully random and probabilistic mental model and the other participants stated the sub-microscopic level was random but displayed mental models that

conflicted with their statement. The participants' conflicting mental models included adding drivers, struggling to conceptualize, or trying to explain randomness, believing that chemical processes are efficient and therefore cannot be random, and believing molecules have a purpose.

When the participants were faced with the simulation, seven of them displayed cognitive dissonance by (1) interpreting the simulation to "fit" their mental models, (2) dismissing the simulation, or (3) using the simulation to change their mental models.

## 4.14 Implications for research

Our research show that participants held multiple mental models and displayed specific mental models in specific contexts. However, we cannot make inferences on why or how they select their mental models. Further insight on how and why a participant selects a mental model can help guide how to teach motion, collisions, and randomness. There is no information on the effect of explicit instructions about randomness, or the effect of representations on the mental models that participants choose to use.

Research that focuses on bridging the connection between mental models and prompts could help create assessments for specific mental models. Since different contexts lead to different mental models, assessment prompts that are aimed at a specific mental model could be researched. This could provide more information about which mental models students possess.

Similarly, further research on the effect of the molecular dynamic visualization on participants' mental models of the sub-microscopic world could provide information on how to best use them in teaching. Previous research has shown that visualizations help students gain dynamic mental models (Kelly and Jones, 2007; Akaygun, 2016; Kelly *et al.*, 2017) and high school students demonstrated big gains using visualization (Ardac and Akaygun, 2004; Adbo and Taber, 2009). However, our results show that for some students the visualization that was shown did not change their mental models. Further research is needed to investigate how and when simulations could be used to provide students with the skills needed to build their mental models. Potentially, scaffolding visualizations into practice as the curriculum progresses could help students gain and build their mental models.

## 4.15 Implications for teaching

Molecular dynamic visualizations can be used during instruction to provide students with the tools to build a mental model of the sub-microscopic level. The molecular dynamic simulations led to some cognitive dissonance that in turn led to changes in mental models. Using simulations in the classroom may help students build the knowledge necessary to visualize the sub-microscopic world, by either confirming their view (like P1) or creating cognitive dissonance that can lead to changes in mental models (like P6, P8, and P10).

Using different prompts and contexts during instruction can provide students with the opportunity to use different mental models. Current assessments tend to focus on the symbolic level, and don't necessarily assess students' mental model of the sub-microscopic level. A robust mental model of chemistry includes understanding the dynamic nature of the sub-microscopic level, therefore using different prompts and contexts (including using molecular dynamic simulation) could provide students with a more scientifically accurate mental model of chemistry.

## 4.16 Limitations

The small sample size (N = 9) means that the findings are exploratory and non-generalizable. This investigation was done at one institution and may not reflect students' thinking at other institutions. Our recruitment was not purposeful and therefore may have some sampling bias. We also recruited from two classes and did not ask students about their levels of study; therefore, students may have different skills and knowledge.

The interviews were designed to elicit specific mental models, without bringing new concepts in. As such, we cannot say with certainty what mental models a participant may hold – we can only infer which mental models were used during the activities in this study. Participants may hold other mental models that were not elicited in this study. We also cannot make inferences on why participants used a specific mental model, or whether the use of a specific mental model was an active choice.

## 4.17 References


Abraham M., Varghese V., and Tang H., (2010), Using Molecular Representations to Aid Student Understanding of Stereochemical Concepts. *Journal of Chemical Education*, **87**(12), 1425–1429.

Adbo K. and Taber K. S., (2009), Learners' Mental Models of the Particle Nature of Matter: A Study of 16-Year-Old Swedish Science Students. *International Journal of Science Education*, **31**(6), 757–786.

Aimeur E., Dufort H., Leibu D., and Frasson C., (1997), Some Justifications About the Learning by Disturbing Strategy, pp. 119–126.

Akaygun S., (2016), Is the Oxygen Atom Static or Dynamic? The effect of Generating Animations on Students' Mental Models of Atomic Structure. *Chemistry Education Research and Practice*, **17**(4), 788–807.

Akaygun S. and Jones L. L., (2013), Animation or Simulation: Investigating the Importance of Interactivity for Learning Solubility Equilibria. *ACS Symposium Series*, **1142**, 127–159.

Alfieri L., Brooks P. J., Aldrich N. J., and Tenenbaum H. R., (2011), Does Discovery-Based Instruction Enhance Learning? *Journal of Educational Psychology*, **103**(1), 1–18.

Ardac D. and Akaygun S., (2004), Effectiveness of Multimedia-Based Instruction That Emphasizes Molecular Representations on Students' Understanding of Chemical Change. *Journal of Research in Science Teaching*, **41**(4), 317–337.

Atkinson M. B., Croisant M., and Bretz S. L., (2021), Investigating First-Year Undergraduate Chemistry Students' Reasoning with Reaction Coordinate Diagrams when Choosing among Particulate-Level Reaction Mechanisms. *Chemistry Education Research and Practice*, **22**(1), 199–213.

Bhattacharyya G. and Bodner G. M., (2005), "It Gets Me to the Product": How Students Propose Organic Mechanisms. *Journal of Chemical Education*, **82**(9), 1402.

Birch P. D. J. and Whitehead A. E., (2019), Investigating the Comparative Suitability of Traditional and Task-Specific Think Aloud Training. *Perceptual and Motor Skills*, **127**(1), 202–224.

Bodner G. M. and Domin D. S., (2000), Mental Models: The Role of Representations in Problem Solving in Chemistry. *University Chemistry Education*, **4**(1), 24–30.

Bongers A., Beauvoir B., Streja N., Northoff G., and Flynn A. B., (2020), Building Mental Models of a Reaction Mechanism: The Influence of Static and Animated Representations, Prior Knowledge, and Spatial Ability. *Chemistry Education Research and Practice*, **21**(2), 496–512.

Bongers A., Northoff G., and Flynn A. B., (2019), Working with Mental Models to Learn and Visualize a New Reaction Mechanism. *Chemistry Education Research and Practice*, **20**(3), 554–569.

Bower G. H. and Morrow D. G., (1990), Mental Models in Narrative Comprehension. *Science*, **247**(4938), 44–48.

Bussey T. J. and Orgill M., (2015), What do Biochemistry Students pay Attention to in External Representations of Protein Translation? The Case of the Shine–Dalgarno Sequence. *Chemistry Education Research and Practice*, **16**(4), 714–730.

Chandrasegaran A. L., Treagust D. F., and Mocerino M., (2007), The Development of a Two-Tier Multiple-Choice Diagnostic Instrument for Evaluating Secondary School Students' Ability to Describe and Explain Chemical Reactions Using Multiple Levels of Representation. *Chemistry Education Research and Practice*, **8**(3), 293–307.

Charmaz K. and Henwood K., (2017), Grounded Theory Methods for Qualitative Psychology, in *The SAGE Handbook of Qualitative Research in Psychology*, pp. 1–23.

Cooper J., (2019), Cognitive Dissonance: Where We've Been and Where We're Going. *International Review of Social Psychology*, **32**(1), 7.

Cooper J. and Worchel S., (1970), Role of Undesired Consequences in Arousing Cognitive dissonance. *Journal of Personality and Social Psychology*, **16**(2), 199–206.

Creswell J. W. and Miller D. L., (2000), Determining Validity in Qualitative Inquiry. *Theory Into Practice*, **39**(3), 124–130.

Dalgarno B., Kennedy G., and Bennett S., (2014), The Impact of Students' Exploration Strategies on Discovery Learning using Computer-Based Simulations. *Educational Media International*, **51**(4), 310–329.

Dangur V., Avargil S., Peskin U., and Dori Y. J., (2014), Learning Quantum Chemistry via a Visual-Conceptual Approach: Students' Bidirectional Textual and Visual Understanding. *Chemistry Education Research and Practice*, **15**(3), 297–310.

DeCocq V. and Bhattacharyya G., (2019), TMI (Too Much Information)! Effects of Given Information on Organic Chemistry Students' Approaches to Solving Mechanism Tasks. *Chemistry Education Research and Practice*, **20**(1), 213–228.

Doerr H. M., (2000), How Can I Find a Pattern in this Random Data? The Convergence of Multiplicative and Probabilistic Reasoning. *Journal of Mathematical Behavior*, **18**(4), 431–454.

Ferguson R. and Bodner G. M., (2008), Making Sense of the Arrow-Pushing Formalism among Chemistry Majors Enrolled in Organic Chemistry. *Chemistry Education Research and Practice*, **9**(2), 102–113.

Festinger L., (1957), *A Theory of Cognitive Dissonance*, Stanford University Press.

Galloway K. R., Leung M. W., and Flynn A. B., (2019), Patterns of Reactions: A Card Sort Task to Investigate Students' Organization of organic chemistry reactions. *Chemistry Education Research and Practice*, **20**(1), 30–52.

Galloway K. R., Stoyanovich C., and Flynn A. B., (2017), Students' Interpretations of Mechanistic Language in Organic Chemistry before Learning Reactions. *Chemistry Education Research and Practice*, **18**(2), 353–374.

Garvin-Doxas K. and Klymkowsky M. W., (2008), Understanding Randomness and its Impact on Student Learning: Lessons Learned from Building the Biology Concept Inventory (BCI). *CBE—Life Sciences Education*, **7**(2), 227–233.

Gibbs G. R., (2018), *Analyzing Qualitative Data*, 2nd ed. SAGE Publication Ltd.

Gilbert J. K., (2008), *Visualization: An Emergent Field of Practice and Enquiry in Science Education*, Gilbert J. K., Reiner M., and Nakleh M. (eds.).

Gove W. R. and Geerken M. R., (1977), Response Bias in Surveys of Mental Health: An Empirical Investigation. *American Journal of Sociology*, **82**(6), 1289–1317.

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

Greca I. M. and Moreira M. A., (2000), Mental Models, Conceptual Models, and Modelling. *International Journal of Science Education*, **22**(1), 1–11.

Günersel A. B. and Fleming S. A., (2013), Qualitative Assessment of a 3D Simulation Program: Faculty, Students, and Bio-organic Reaction Animations. *Journal of Chemical Education*, **90**(8), 988–994.

Heemsoth T. and Heinze A., (2014), The Impact of Incorrect Examples on Learning Fractions: A Field Experiment with 6th Grade Students. *Instructional Science*, **42**(4), 639–657.

Höffler T. N. and Leutner D., (2007), Instructional Animation Versus Static Pictures: A meta-Analysis. *Learning and Instruction*, **17**(6), 722–738.

Johnson-Laird P. N., (2010), Mental Models and Human Reasoning. *Proceedings of the National Academy of Sciences*, **107**(43), 18243–18250.

Johnstone A. H., (1991), Why is Science Difficult to Learn? Things are Seldom what they Seem. *Journal of Computer Assisted Learning*, **7**(2), 75–83.

Jolfaee S., Zazkis R., and Sinclair N., (2014), "It Is Very, Very Random Because It Doesn't Happen Very Often": Examining Learners' Discourse on Randomness BT - Probabilistic Thinking: Presenting Plural Perspectives, in Chernoff E. J. and Sriraman B. (eds.)., Springer Netherlands, pp. 397–416.

Jones L. L. and Kelly R. M., (2015), Visualization: The Key to Understanding Chemistry Concepts, in *ACS Symposium Series*, Orna M. V. (ed.)., American Chemical Society, pp. 121–140.

Kelly R. and Hansen S. J. R., (2017), Exploring the Design and Use of Molecular Animations that Conflict for Understanding Chemical Reactions. *Química Nova*.

Kelly R. M., (2014a), Using Variation Theory with Metacognitive Monitoring to Develop Insights into how Students Learn from Molecular Visualizations. *Journal of Chemical Education*, **91**(8), 1152–1161.

Kelly R. M. and Akaygun S., (2016), Insights into How Students Learn the Difference between a Weak Acid and a Strong Acid from Cartoon Tutorials Employing Visualizations. *Journal of Chemical Education*, **93**(6), 1010–1019.

Kelly R. M. and Akaygun S., (2019), Visualizations and Representations in Chemistry Education. *Chemistry Education Research and Practice*, **20**(4), 657–658.

Kelly R. M., Akaygun S., Hansen S. J. R., and Villalta-Cerdas A., (2017), The Effect that Comparing Molecular Animations of Varying Accuracy has on Students' Submicroscopic Explanations. *Chemistry Education Research and Practice*, **18**, 582–600.

Kelly R. M., Akaygun S., Hansen S. J. R., Villalta-Cerdas A., and Adam J., (2021), Examining Learning of Atomic Level Ideas About Precipitation Reactions with a Resources Framework. *Chemistry Education Research and Practice*, **22**(4), 886–904.

Kelly R. M. and Jones L. L., (2007a), Exploring How Different Features of Animations of Sodium Chloride Dissolution Affect Students' Explanations. *Journal of Science Education and Technology*, **16**(5), 413–429.

Kelly R. M. and Jones L. L., (2007b), Exploring how Different Features of Animations of Sodium Chloride Dissolution affect Students' Explanations. *Journal of Science Education and Technology*, **16**(5), 413–429.

Kelly R. M. and Jones L. L., (2008), Investigating Students' Ability to Transfer Ideas Learned from Molecular Animations of the Dissolution Process. *Journal of Chemical Education*, **85**(2), 303.

Kirschner P. A., Sweller J., and Clark R. E., (2006), Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, **41**(2), 75–86.

Kozma R. B. and Russell J., (1997), Multimedia and Understanding: Expert and Novice Responses to Different Representations of Chemical Phenomena. *Journal of Research in Science Teaching*, **34**(9), 949–968.

Krippendorff K., (2004), Reliability in Content Analysis.: Some Common Misconceptions and Recommendations. *Human Communication Research*, **30**(3), 411–433.

Linder D. E., Cooper J., and Jones E. E., (1967), Decision Freedom as a Determinant of the Role of Incentive Magnitude in Attitude Change. *Journal of Personality and Social Psychology*, **6**(3), 245–254.

Luisi P. L., (2002), Emergence in Chemistry: Chemistry as the Embodiment of Emergence. *Foundations of Chemistry*, **4**(3), 183–200.

Macrie-Shuck M. and Talanquer V., (2020), Exploring Students' Explanations of Energy Transfer and Transformation. *Journal of Chemical Education*, **97**(12), 4225–4234.

Mayer R. E., (2005), *The Cambridge Handbook of Multimedia Learning*, Cambridge University Press.

Mayer R. E. and Estrella G., (2014), Benefits of Emotional Design in Multimedia Instruction. *Learning and Instruction*, **33**, 12–18.

McCambridge J., de Bruin M., and Witton J., (2012), The Effects of Demand Characteristics on Research Participant Behaviors in Non-Laboratory Settings: A Systematic Review. *PLoS ONE*, **7**(6), e39116.

McHugh M. L., (2012), Interrater Reliability: The Kappa Statistic. *Biochemia medica*, **22**(3), 276–282.

Miller G. A., (1956), The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. *Psychological Review*, **63**(2), 81–97.

Moore E. B., Chamberlain J. M., Parson R., and Perkins K. K., (2014), PhET Interactive Simulations: Transformative Tools for Teaching Chemistry. *Journal of Chemical Education*, **91**(8), 1191–1197.

Moore E. B., Herzog T. A., and Perkins K. K., (2013), Interactive Simulations as Implicit Support for Guided-Inquiry. *Chemistry Education Research and Practice*, **14**(3), 257–268.

Ollino M., Aldoney J., Domínguez A. M., and Merino C., (2018), A new Multimedia Application for Teaching and Learning Chemical Equilibrium. *Chemistry Education Research and Practice*, **19**(1), 364–374.

Popova M. and Bretz S. L., (2018), It's Only the Major Product that we Care About in Organic Chemistry: An Analysis of Students' Annotations of Reaction Coordinate Diagrams. *Journal of Chemical Education*, **95**, in press.

Rapp D. N., (2005), Mental models: Theoretical Issues for Visualizations in Science Education, in *Visualization in science education*, pp. 43–60.

Rapp D. N. and Kurby C. A., (2008), The 'Ins' and 'Outs' of Learning: Internal Representations and External Visualizations, in *Visualization: An Emergent Field of Practice and Enquiry in Science Education*, pp. 29–52.

Rosenthal D. P. and Sanger M. J., (2013), How Does Viewing One Computer Animation Affect Students' Interpretations of Another Animation Depicting the Same Oxidation–Reduction Reaction? *Chemistry Education Research and Practice*, **14**(3), 286–296.

Rosenthal D. P. and Sanger M. J., (2012), Student Misinterpretations and Misconceptions Based on their Explanations of two Computer Animations of Varying Complexity Depicting the Same Oxidation–Reduction Reaction. *Chemistry Education Research and Practice*, **13**(4), 471–483.

Sanger M. J., Campbell E., Felker J., and Spencer C., (2007), "Concept Learning versus Problem Solving": Does Particle Motion Have an Effect? *Journal of Chemical Education*, **84**(5), 875.

Savard A., (2014), Developing Probabilistic Thinking: What About People's Conceptions? BT - Probabilistic Thinking: Presenting Plural Perspectives, in Chernoff E. J. and Sriraman B. (eds.)., Springer Netherlands, pp. 283–298.

Stieff M. and Wilensky U., (2003a), Connected Chemistry—Incorporating Interactive Simulations into the Chemistry Classroom. *Journal of Science Education and Technology*, **12**(3), 285–302.

Strickland A. M., Kraft A., and Bhattacharyya G., (2010), What Happens when Representations Fail to Represent? Graduate Students' Mental Models of Organic Chemistry Diagrams. *Chemistry Education Research and Practice*, **11**(4), 293–301.

Suits J. P., (2015), Design of Dynamic Visualizations to Enhance Conceptual Understanding in Chemistry Courses. *Chemistry Education: Best Practices, Opportunities and Trends*, 595–620.

Sweeder R. D., Herrington D. G., and VandenPlas J. R., (2019), Supporting Students' Conceptual Understanding of Kinetics using Screencasts and Simulations Outside of the Classroom. *Chemistry Education Research and Practice*, **20**(4), 685–698.

Taber K. S., (2009), Learning at the Symbolic Level, in *Multiple Representations in Chemical Education*, Gilbert J. K. and Treagust D. (eds.) Models and Modeling in Science Education. Springer Netherlands, pp. 75–105.

Taber K. S. and García-Franco A., (2010), Learning Processes in Chemistry: Drawing Upon Cognitive Resources to Learn About the Particulate Structure of Matter. *Journal of the Learning Sciences*, **19**(1), 99–142.

Talanquer V., (2007), Explanations and Teleology in Chemistry Education. *International Journal of Science Education*, **29**(7), 853–870.

Talanquer V., (2011), Macro, Submicro, and Symbolic: The many faces of the chemistry "triplet." *International Journal of Science Education*, **33**(2), 179–195.

Talanquer V., (2008), Students' Predictions About the Sensory Properties of Chemical Compounds: Additive Versus Emergent Frameworks. *Science Education*, **92**(1), 96–114.

Talanquer V., (2015), Threshold Concepts in Chemistry: The Critical Role of Implicit Schemas. *Journal of Chemical Education*, **92**(1), 3–9.

Talanquer V., (2013), When Atoms Want. *Journal of Chemical Education*, **90**(11), 1419–1424.

Tang H. and Abraham M. R., (2016), Effect of Computer Simulations at the Particulate and Macroscopic Levels on Students' Understanding of the Particulate Nature of Matter. *Journal of Chemical Education*, **93**(1), 31–38.

Tasker R. and Dalton R., (2006a), Research into Practice: Visualisation of the Molecular World using Animations. *Chemistry Education Research and Practice*, **7**(2), 141.

Tasker R. and Dalton R., (2008), Visualizing the Molecular World - Design, Evaluation, and Use of Animations. *Visualization Theory and Practice in Science Education*, **3**, 103–131.

Teichert M. A., Tien L. T., Anthony S., and Rickey D., (2008), Effects of Context on Students' Molecular-Level Ideas. *International Journal of Science Education*, **30**(8), 1095–1114.

Tuvi-Arad I. and Gorsky P., (2007), New Visualization Tools for Learning Molecular Symmetry: A Preliminary Evaluation. *Chemistry Education Research and Practice*, **8**(1), 61–72.

Tversky B., Morrison J. B., and Betrancourt M., (2002), Animation: Can it Facilitate? The Enthusiasm for Graphics of all Kinds Rests on the Belief that they Benefit Comprehension

and Learning, and Foster Insight. *Int. J. Human-Computer Studies Schnotz & Kulhavy*, **57**, 247–262.

Uttal D. H. and O'Dogerty K., (2008), Comprehending and Learning from 'Visualizations': A Developmental Perspective, in *Visualization: An Emergent Field of Practice and Enquiry in Science Education*, pp. 53–72.

Velázquez-Marcano A., Williamson V. M., Ashkenazi G., Tasker R., and Williamson K. C., (2004), The Use of Video Demonstrations and Particulate Animation in General Chemistry. *Journal of Science Education and Technology*, **13**(3), 315–323.

Venkataraman B., (2009), Visualization and Interactivity in the Teaching of Chemistry to Science and Non-Science Students. *Chemistry Education Research and Practice*, **10**(1), 62–69.

Watson S. W., Dubrovskiy A. V., and Peters M. L., (2020), Increasing Chemistry Students' Knowledge, Confidence, and Conceptual Understanding of pH using a Collaborative Computer pH Simulation. *Chemistry Education Research and Practice*, **21**(2), 528–535.

Wavefunction, (2019), *Odyssey*, Wavefunction.

Williamson V. M. and Abraham M. R., (1995), The Effects of Computer Animation on the Particulate Mental Models of College Chemistry Students. *Journal of Research in Science Teaching*, **32**(5), 521–534.

Williamson V. M., Lane S. M., Gilbreath T., Tasker R., Ashkenazi G., Williamson K. C., and Macfarlane R. D., (2012), The Effect of Viewing Order of Macroscopic and Particulate Visualizations on Students' Particulate Explanations. *Journal of Chemical Education*, **89**, 979–987.

Wright L., (2020), *Teleological explanations*, University of California Press.

Yan F. and Talanquer V., (2015), Students' Ideas about How and Why Chemical Reactions Happen: Mapping the conceptual landscape. *International Journal of Science Education*, **37**(18), 3066–3092.

Yezierski E. J. and Birk J. P., (2006), Misconceptions about the Particulate Nature of Matter. Using Animations to Close the Gender Gap. *Journal of Chemical Education*, **83**(6), 954–960.

# Chapter 5. OrgChem101.com: Mastering the arrows online module: what are the learning gains after using the module?

*"We need to find a way to help good problem-solvers develop into good chemists. Within the context of organic mechanisms, this means helping the students recognize that there are concepts and principles that lie behind each step in the mechanisms they propose that need to be more clearly comprehended." (Bhattacharyya, 2005)*

## 5.1 Context: Student's struggle to use the electron pushing formalism to predict reaction

The electron-pushing formalism (EPF) is a representation of chemistry reactivity used to explain sub-microscopic phenomenon. EPF is used extensively by experts as a tool to predict reactions, explain sub-microscopic level phenomenon, and explain specific reactivity (Bhattacharyya, 2013). However, there have been many reports that learners do not use this tool as intended (Bhattacharyya and Bodner, 2005; Ferguson and Bodner, 2008; Grove, Cooper, and Cox, 2012; Bhattacharyya, 2013; Graulich, 2015; DeCocq and Bhattacharyya, 2019). Learners may not see the value of using EPF and instead use rote memorization to solve chemistry problems (Ferguson and Bodner, 2008). To help learners use EPF as a predictive tool, the University of Ottawa introduced a new Organic Chemistry curriculum (Flynn and Ogilvie, 2015), along with an online module for learning (Flynn et al., 2016). In this chapter, I will describe the investigation of the effect of the online module *OrgChem101.com – Mastering the arrows* on students' skills and their approach to solving EPF problems.

## 5.2 Literature review: symbolic language of organic chemistry – the electron pushing formalism

The EPF symbols are curved arrows that represent electron movement. The curved arrow starts at an electron pair, either a lone pair or a double bond, and points to the location that the electrons are moving to, indicating bond breaking and forming events. For example, in Figure 5.1 the blue arrow starts on a lone pair of electrons on the oxygen atom and points to the carbon atom on the carbonyl; this arrow represents the movement of electron from the

hydroxide ($OH^-$) to the carbon atom on the carbonyl. As such, the EPF arrows are used in organic chemistry for predicting reactions and understanding how electrons behave (Bhattacharyya, 2013)

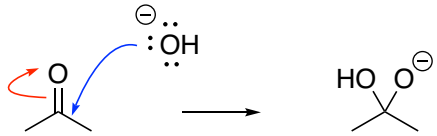



*Figure 5.1. An example of EPF using hydroxide and a carbonyl.*

Several reports show that learners do not use EPF as intended (i.e., a tool to solve problems and predict a reaction pathway). Mastery of EPF has been investigated in many different studies and there are multiple literature reports in the literature of students (1) using EPF as an afterthought and not as a tool to predict reactions, (2) relying on rote memorization to solve problems, (3) using deterministic thinking, or (4) reasoning about a single factor.

Learners often use EPF as an after-thought once a product has been determined (Bhattacharyya and Bodner, 2005; Grove, Cooper, and Rush, 2012; Bhattacharyya, 2013; Caspari et al., 2017; DeCocq and Bhattacharyya, 2019). Many learners will begin solving a problem by mapping the differences between products and reactants in order to recall EPF or specific steps (Bhattacharyya and Bodner, 2005; Bhattacharyya, 2014). When Bhattacharyya and Bodner (2005) asked participants in their study "*Why did you propose this step?",* many learners answered: "*Because it gets me the product*". This type of reasoning often led learners to propose EPF that lead of highly unstable products, or ones that were not feasible (Bhattacharyya and Bodner, 2005). When given an entire transformation (i.e., reagents and products), learners often reverted to backward reasoning (e.g., using the product to propose the arrows); however, when the product was omitted, participants used EPF as a predictive tool more readily (DeCocq and Bhattacharyya, 2019).

Learners often relied on rote memorization of reactions as opposed to using EPF as a tool to predict reactions (Ferguson and Bodner, 2008). Using EPF to predict reactions requires a high cognitive demand and instead revert to rote memorization (Grove, Cooper, and Cox, 2012). In a similar vein, learners sometimes use intuition and heuristics (i.e., rules of thumb or shortcuts) to find a product rather than using EPF to predict the reaction (Graulich, 2015).

Deterministic thinking occurs when a person believes that "chemical processes are believed to be driven by specific forces to yield a specific product" (Bhattacharyya, 2014). This type of reasoning will often lead to learners not considering alternative reaction pathways (Caspari *et al.*, 2017; Lieber and Graulich, 2020). Educators will often talk about EPF arrows as "productive"; however, it is unclear how learners may interpret the word. While the educators understand that productive arrows are a possible path and that many possible alternate arrows exist, learners may interpret the word "productive" as deterministic (Caspari *et al.*, 2017).

When using EPF to predict reactions multiple variables are involved. Multi-variate reasoning occurs when multiple variables will affect an outcome and those variables must be weighed against each other (Kraft *et al.*, 2010). First one must determine nucleophilic and electrophilic sites in molecules, which has been shown to be difficult for many learners (Anzovino and Bretz, 2015). To identify nucleophilic and electrophilic sites, multiple factors must be considered such as electronegativity of atoms, inductive effect, delocalization of electrons, steric hinderance, etc., and then a reaction pathway must be determined as feasible or not. Fourth-year learners in a study were often confused about which factors were important to consider (Rushton *et al.*, 2008). Other learners would rely on isolated featured, as opposed to consider all the parts of a reaction to predict their pathways (DeFever *et al.*, 2015). While our study did not address this type of reasoning, it is important to note that predicting reactions involves multiple factors, and that we cannot pinpoint exactly how students weigh those factors.

## 5.3 University of Ottawa Organic Chemistry Curriculum: Electron Pushing Formalism first and patterns of reactivity

To address the problem of rote memorization, the University of Ottawa introduced a new curriculum called "Mechanistic patterns and principles" (referenced in this thesis as the *Mechanistic Patterns curriculum)* in 2012 which focuses on mechanistic patterns and introduced EPF first (Flynn and Ogilvie, 2015). This curriculum differs from other organic chemistry curricula, which often teach by surface features (i.e., functional groups). In the *Mechanistic Patterns curriculum* learners learn the symbolism (i.e., EPF) before any reactions. This sequence provides the learners with the skills to become fluent in the language of EPF. Learning EPF first, provides the learners with the skills to understand what the representation means and become fluent in

the language of EPF. Several research studies have investigated how the *Mechanistic Patterns curriculum* impacted students' skills using EPF and knowledge of reactivity principles (Bodé and Flynn, 2016; Flynn and Featherstone, 2017; Galloway, Stoyanovich, and Flynn, 2017; Webber and Flynn, 2018).

The qualitative investigation of the effect of the *Mechanistic Patterns curriculum* have shown that learners scored higher on mechanistic questions when enrolled in the Mechanistic *Patterns curriculum* as opposed to the non-*Mechanistic Patterns curriculum* (Webber and Flynn, 2018). In think aloud interviews, Webber and Flynn (2018) found that learners the *Mechanistic Patterns curriculum* were skillful at proposing curved arrows and that only one participant used backward thinking. Unprompted, the learners used several problem-solving strategies such as expanding the structure, identifying nucleophiles and electrophiles, mapping, and making multiple attempts. Students in the study also used many underlying concepts in their answers and showed some causal reasoning. Other participants would split an elementary step as a potential form of cognitive offloading. This study showed that students in the *Mechanistic Patterns curriculum* could predict reactions and used the curved arrows and underlying principles to solve mechanism problems.

In a larger scale study, 33 organic chemistry exam questions were analyzed; 15 questions asked students to propose arrows for a mechanism and 17 questions asked students to propose the products from arrows drawn (Flynn and Featherstone, 2017). Student answers were mostly correct and did not include many erroneous arrows (e.g., reversed arrows; Flynn and Featherstone, 2017). Flynn and Featherstone (2017) also found that different questions had different success rates and that questions about implicit atoms, intramolecular reactions, and the conformation of the reactants were correlated with lower success rates. These types of errors could be related to visualization of the molecules (i.e., the reactions required mental rotations), as opposed to EPF skills.

In a different study, participants were recruited from courses following the *Patterns* curriculum ($N$ = 11) and the participants used electron movement to explain mechanisms and curved arrows (Galloway, et.al., 2017). When answering draw the product questions some participants would rely on charge (i.e., an explicit feature) to answer the problem. Expanding on

this study, a card sort instrument was developed and implemented to investigate if students used surfaces features or mechanistic patterns to solve problems (Galloway et al., 2018; 2019; Lapierre and Flynn, 2019; Lapierre et al., 2022). The card-sort studies found that as students progressed through the Organic Chemistry sequence, they became more successful at sorting reactions by mechanism patterns (Lapierre and Flynn, 2019). Student who sorted reactions by mechanistic patterns on early low-stake assessments had higher scores on their final exams (Lapierre et al., 2022).

Experts can see patterns in reaction mechanisms (i.e., curved arrows in a reaction) and learners following the *Mechanistic Patterns curriculum* can start sorting reactions like experts (Lapierre, 2019; Galloway et.al., 2018; 2019). Students in the *Mechanistic Patterns curriculum* are skillful at proposing curved arrows (Flynn and Featherston, 2017; Webber and Flynn, 2018), using arrows to represent electron movement (Galloway, et.al., 2017), and mapping atoms and electrons (Bodé and Flynn, 2016; Galloway, et.al., 2017; Webber and Flynn, 2017; Flynn and Featherstone, 2018). Teaching students the language of organic chemistry (e.g., EPF) could provide the tools to students to think like experts.

## 5.4 Learning tool: OrgChem101.com – Mastering the Arrows

To support the *Mechanistic Patterns curriculum*, a team at the University of Ottawa designed an online website (OrgChem101.com) that aligns with the curriculum. The website that contains three modules: one for nomenclature (Flynn *et al.*, 2014), one for acid-base chemistry (Stoyanovich *et al.*, 2015) and one for the symbolism of reaction mechanisms. While the effects of two of the modules have been evaluated (Stoyanovich *et al.*, 2015; Bodé *et al.*, 2016), the effects of the mechanism modules on student learning have not been evaluated.

"OrgChem101.com – Mastering the Arrows" is a free bilingual (EN/FR) online module to help students learn about EPF (Flynn and Visser, 2018). Since cost and access can be a barrier to students, this module was created as an Open Education Resources (OER) and is available to anyone who can access the internet. The module is designed to be student-centered; the learners decide how to learn and can learn at their own pace (Mayer *et al.*, 2001; Hannafin, 2012). The module also allows students to practice the concept and provides feedback. The module also includes a metacognitive component, where learners are prompted to rate their skills on LOs.

The goal of the module is to teach learners how to use EPF for reactions they have never seen before, therefore requiring them to use EPF as a predictive tool. The module is focused on four LOs:

- Draw the electron pushing arrows
- Draw the product of a reaction step
- Draw the transition state of a reaction step
- Draw the mechanism and electron pushing arrows of the reverse reaction

A concept is first introduced via an interactive video that the student controls. Then a series of practice questions (with feedback) is available. All the questions and videos are for new reactions that most learners in Organic Chemistry I at the University of Ottawa would not have encountered. Since the reactions would not have been memorized, learners must use the EPF arrows to predict the reaction. The module also teaches a variety of strategies for solving EPF questions.

## 5.5 Goals and Research Questions

The goal of this research was to evaluate the online Organic Mechanisms module and to determine the learning gains for learners. The hypothesis for this research was that that learners using the module would have more success achieving the LOs and would be able to propose curved arrows for new reactions. Since the module also teaches strategies for problem solving, we also hypothesized that the module would affect the use of those strategies as well as the types of errors learners may make.

**RQ1**: What are learners' learning gains after using the Organic Mechanisms module?

**RQ2**: What effect does the Organic Mechanisms module have on strategies used and errors made when learners solve EPF-related questions?

## 5.6 Published contribution

*Evaluating students' learning gains, strategies, and errors using OrgChem101's Module: Organic Mechanisms — Mastering the arrows*

Myriam S. Carle, Rebecca Visser, and Alison B. Flynn

**Reproduced with permission from**: Carle, M. S., Visser, R., & Flynn, A. B. (2020). Evaluating students' learning gains, strategies, and errors using OrgChem101's module: Organic mechanisms—mastering the arrows. *Chemistry Education Research and Practice*, *21*(2), 582–596. https://doi.org/10.1039/C9RP00274J

**Abstract:** We developed an online learning module called "Organic Mechanisms: Mastering the Arrows" to help students learn part of organic chemistry's language—the electron-pushing formalism. The module guides students to learn and practice the electron-pushing formalism using a combination of interactive videos, questions with instant feedback, and metacognitive skill-building opportunities. This module is part of OrgChem101.com, an open educational resource (OER) that houses a series of learning modules. To evaluate the mechanism module's effects on students' learning and experiences, we offered a workshop during which undergraduate students used the module. We investigated their learning gains via a pre-test and post-test format and their experiences using a survey. Analysis of responses revealed significant learning gains between the pre- and post-test, especially with questions that asked students to draw the products of a reaction. After using the learning tool, students used more analysis strategies, such as mapping, attempted more questions, and made fewer errors. The students reported positive experiences and a belief that the module would help them in their organic chemistry courses. Previous work also identified greater metacognitive skills after using the module, related to the module's intended learning outcomes. Herein, we describe the module, evaluation study, findings, and implications for research and practiced.

**Author contribution:** RV designed the workshops and original assessment and did data collection and analysis for the pilot studies. Some modifications to the original study were done by MSC to take the study to full scale. MSC did the data collection, data analysis for the full-scale

study, drafted the manuscript, and edited the manuscript based on feedback. The authors discussed all phases of the work. ABF provided guidance on all phases.

## 5.7 Introduction

### 5.7.1 Translating between the molecular and representational levels: chemistry's language

A cohesive understanding in science, technology, engineering, and mathematics (STEM) relies on making sense of processes that are often invisible, operate at many scales (*e.g.*, size), and are depicted with specialized representations (Johnstone, 1982, 2000; Kozma and Russell, 1997; Gilbert, 2008; Talanquer, 2011). Learners often do not know which aspects of these representations are significant, much less how to develop their mental models, translate between representations, and apply their knowledge to solve complex issues (Johnstone, 2000; Bhattacharyya and Bodner, 2005; Russ *et al.*, 2008; Cheng and Gilbert, 2009; Gilbert and Treagust, 2009; Grove, Cooper, and Rush, 2012; National Research Council, 2012; National Research Council (NRC), 2012; Weinrich and Talanquer, 2015; Becker *et al.*, 2016; Cooper *et al.*, 2016; Caspari *et al.*, 2017; Weinrich and Sevian, 2017; Bodé and Flynn, 2018; Moreira *et al.*, 2018; Vasilyeva *et al.*, 2018), akin to working in a second language in which many symbols, representations, and tools function as STEM's words, grammar, and syntax (Taber, 2009; Taskin and Bernholt, 2014). The electron-pushing formalism (EPF) forms part of organic chemistry's language, using curved arrows to represent the electron flow in reaction mechanism (Figure 1). These curved arrows start with electrons (non-bonding electrons or a bond) and point to an electron deficient atom. Despite their use as a tool for experts (Bhattacharyya, 2013), students encounter a number of challenges using the EPF (Ferguson and Bodner, 2008; Craig *et al.*, 2012; Grove and Bretz, 2012; Bhattacharyya, 2014; Anzovino and Bretz, 2015; Graulich, 2015).

To support student mastery of and fluency in the EPF, we created the "Organic Mechanisms: Mastering the Arrows" module in OrgChem101.com (Flynn and Visser, 2018). The module addresses four learning outcomes (LOs, Figure 1): (1) Draw the electron-pushing arrows, given the starting materials and products of the elementary steps, (2) Draw the products of a reaction step, given the starting materials and electron-pushing arrows, (3) Draw the transition state structure for a reaction step, and (4) Draw the reverse reaction mechanism, given the elementary steps in the forward direction (Flynn and Ogilvie, 2015). These LOs are designed to

help students gain fluency that they can leverage when they are learning more advanced concepts of reactivity.

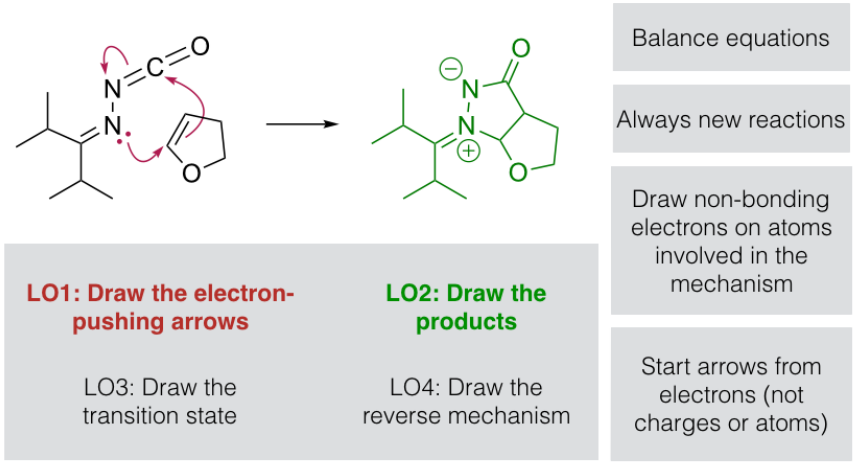



*Figure 5.2. LOs and key principles associated with the electron-pushing formalism*

The organic reaction mechanisms module begins with a "Get started" section that has a self-assessment as part of metacognitive skill-building (Brown *et al.*, 2014), pre-test, a comparison of those two assessments with a prompt that asks students to decide what to do next for their learning, and an introductory video. Each LO section involves interactive videos and activities with feedback. The module finishes with a "Wrap-up" section that has a self-assessment, post-test, and summary. All the sections are aligned with the module's intended LOs and the module can be used in any curriculum type, including traditional and transformed (Flynn and Ogilvie, 2015).

The module introduces strategies for analyzing questions, such as: expanding or redrawing the structure, mapping the atoms and electrons involved in the reaction step, and building a model. Expanding the structure includes drawing all non-bonding electrons (lone pairs) on heteroatoms involved in the reaction steps. Expanding the structure also helps avoid errors related to implicit atoms and electrons, including avoiding pentavalent carbon atoms. Mapping involves keeping track of electrons and atoms from the starting materials to the products, which usually involves numbering or lettering atoms and electrons, but which can take other forms, such as circling atoms, writing geometrical shapes, or highlighting atoms or electrons. Mapping is particularly useful to determine the bonds broken and formed during a reaction step and can

help students avoid mistakes such as missing atoms or misplaced atoms. Building a model can be used to examine the molecule in various conformations, including the reactive one, plus drawing products that have more complex stereochemical information. More details about the module are available in the module itself and in an earlier publication (Visser and Flynn, 2018).

The module was designed based on existing literature, including a study that used exam analysis of thousands of isolated EPF questions that found that students who were taught and assessed on those LOs used few reversed arrows (*e.g.*, atom to electrons), few pentavalent atoms, and higher scores for *Draw arrows* (LO1) than *Draw products* (LO2); lower scores were correlated with questions involving implicit atoms and electrons, intramolecular reaction steps, and reactants drawn in conformations differing the reactive one (Flynn and Featherstone, 2017). A follow-up study used an interview format to investigate students' meaning making when analyzing EPF questions, finding that all participants analyzed electron movement and leveraged their prior knowledge while approaching these questions and the most successful students used mapping strategies (Galloway, Stoyanovich, and Flynn, 2017). Participants relied on charge as a cue to identify areas of reactivity and some used stepwise approaches that resulted in non-chemically feasible intermediates—the latter approach may have simply been a problem-solving strategy to reduce cognitive load or may represent how participants visualized the reactions occurring. Expanding and mapping strategies have also been correlated with successful problem-solving in organic synthesis (Bodé and Flynn, 2016).

### 5.7.2 OrgChem101.com and the modules' impacts

The Organic Mechanisms module (Flynn *et al.*, 2016) forms part of a larger suite of modules hosted on OrgChem101.com: (1) the Nomenclature101 Module, which helps students learn the International Union of Pure and Applied Chemistry (IUPAC) nomenclature and functional groups through 10-question quizzes (Flynn *et al.*, 2014), (2) the Mechanisms Module which focuses on teaching EPF (Visser and Flynn, 2018), and (3) the acid–base module, which is aligned with acid–base LO (Stoyanovich *et al.*, 2015). Each module is student-controlled, interactive, available in English and French, and is an open education resource (free and open access). The latter two modules have metacognitive skill-building layers, designed to help students identify what they know, need to know, and plan their learning strategies. We

previously conducted an educational evaluation to determine the Nomenclature101 module's effectiveness for learning and students' experiences using the learning tool (Bodé *et al.*, 2016). Using a pre-/post-test format and questionnaires, we found high learning gains in a short period of time, plus high student satisfaction. The other modules have not yet been evaluated with respect to their impacts on student learning and experiences, although students' accuracy estimating their ability (an aspect of metacognition) increased after using the module (Visser and Flynn, 2018). In the present study, we examine students' learning and experiences when using the Organic Mechanisms Module (Figure 2).

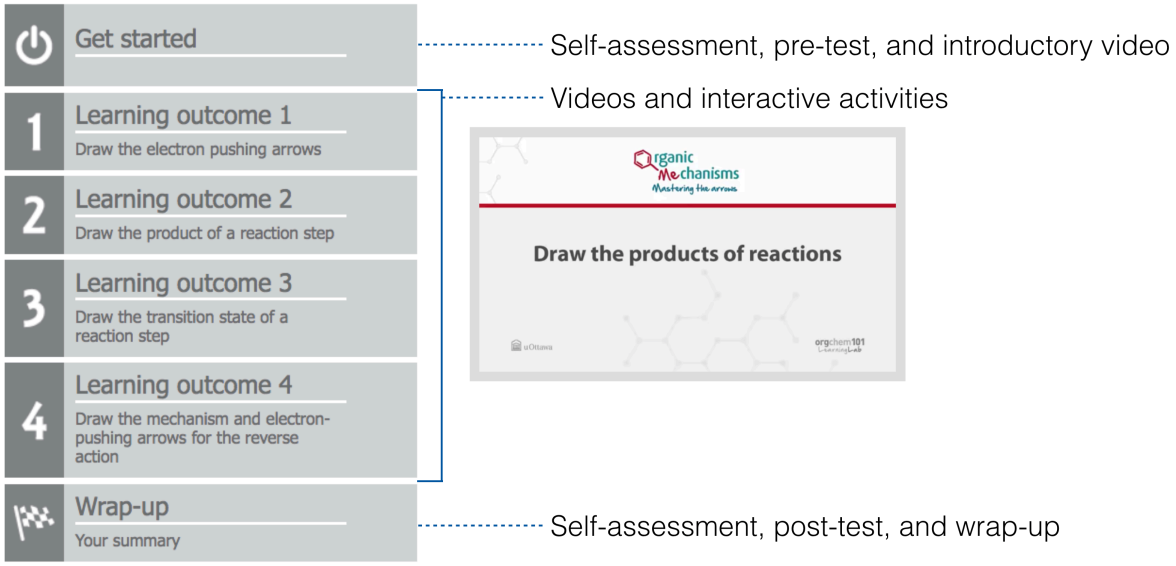



*Figure 8. Overview of the website.*

## 5.8 Theoretical framework

The module's design and evaluation was guided by information processing theory (IPT) (Ausubel *et al.*, 1978; Sweller, 1999; Kalyuga *et al.*, 2003; Mayer and Moreno, 2003; Schunk, 2016). IPT states that people perceive information through their senses and parse the new information in working memory. The learners then integrate information into long-term memory by integrating the information with relevant prior knowledge. The learner must actively choose to integrate the information by perceiving it as meaningful and linking it to prior knowledge (Ausubel *et al.*, 1978; Novak, 1993; Bretz, 2001; Galloway and Bretz, 2015; Schunk, 2016). For this

reason, new knowledge must be presented in a meaningful way to the learner (Novak, 1993). Modern IPT also incorporates the learner's affect and the learning environment (Schunk, 2016).

IPT guided the design of the module; information is presented to student in a meaningful manner by linking the new concepts to what learners have previously learned and to future required (tested) skills. The module is interactive and engages the learner by pausing the videos and allowing the students to build their own answers. The module also contains practice questions with feedback, allowing the student to practice the skills learned and build on their knowledge. Metacognitive skill-building activities and prompts (e.g., compare your skill rating with your pre-test score: how will you manage your studying accordingly?) provide opportunities for learners to increase their skill in identifying what they know and don't know, as well as planning their learning time accordingly (Council, 2000).

## 5.9 Research questions

We focused on three research questions regarding undergraduate students' learning and experiences when using the Organic Mechanisms module in Orgchem101.com.

**RQ1:** What are students' learning gains after using the module?

**RQ2:** What effect does the module have on students' strategies when solving EPF-related questions?

**RQ3:** What effect does the module have on students' errors when solving EPF-related questions?

## 5.10 Methods

### 5.10.1 Setting and course

Participants in the study were taking Organic Chemistry I (cohort I study) and II (cohort II and cohort III studies) courses at a large, research-intensive Canadian university. 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 Chemistry I course has a required, associated laboratory section

(3 hours biweekly) and the Organic Chemistry II course has a laboratory course that runs concurrently and is only required for some programs (3 hours weekly). The University of Ottawa uses a principles and patterns of mechanisms curriculum; in that curriculum, the electron-pushing formalism is explicitly taught before deeper concepts of reactivity are addressed (Flynn and Ogilvie, 2015).

### 5.10.2 Participants and the study's structure

The University of Ottawa's Research Ethics Board approved all phases of this study. Cohort I study informed the design of the Cohort II study described herein and are described in previously published work (Visser and Flynn, 2018) and in Appendix 2.1. The cohort I studies consisted of four parts (Figure 5.3): (1) a pre-test, (2) time allotted for the students to use the learning module, (3) a post-test, and (4) a survey asking for students' opinions and feelings about the module (a more in-depth description is available in Appendix 2.1). The workshop focused on the first two LOs (*Draw the arrows* and *draw the products*).

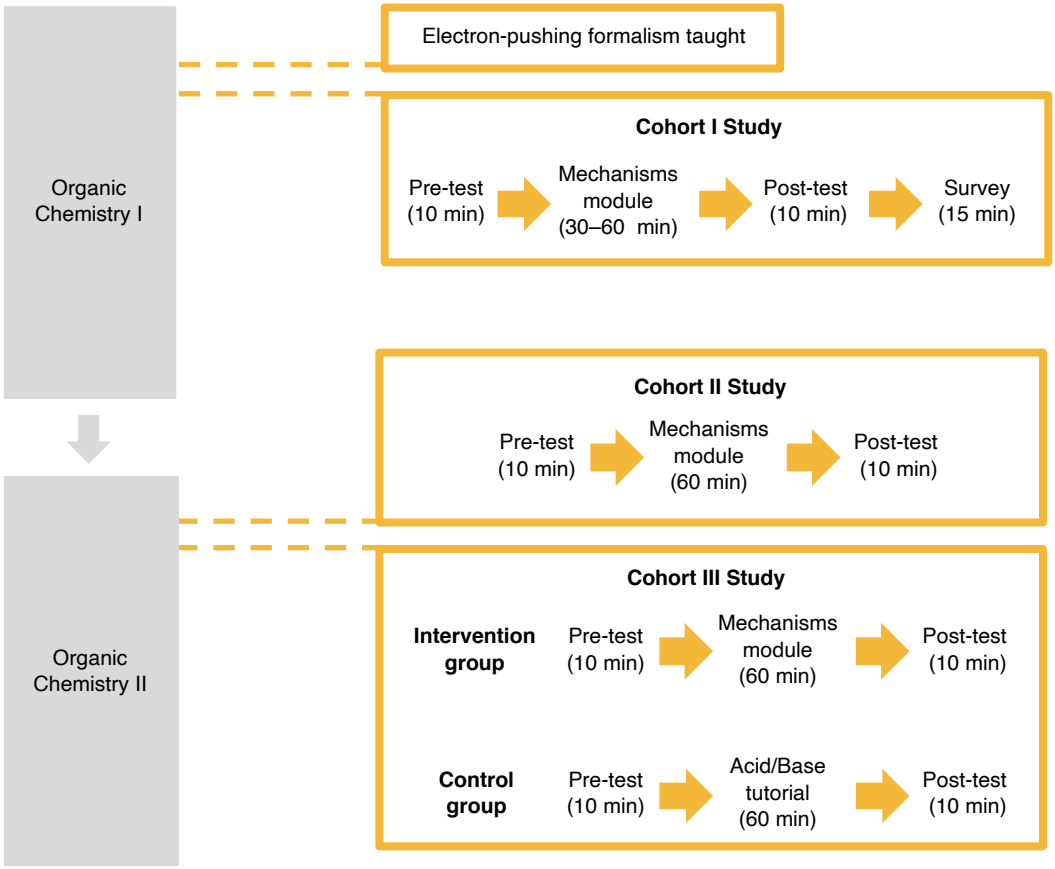



*Figure 5.3. Overview of the studies.*

For the cohort II study described herein, Organic Chemistry II students enrolled in the 2018 fall term were invited to participate in a workshop held during a regularly held tutorial session (*i.e.*, recitation, discussion group). The researchers made an announcement during a class period and the professor teaching the course posted a recruitment text on the class's online page. Workshop attendees provided informed consent to participate in the study and could elect to participate in the workshop without having their data used for the study; 103 of 172 attendees consented to have their data used for research purposes; 330 students were enrolled in the course in total.

The cohort III study consisted of Organic Chemistry II students enrolled in the Fall 2019 term during their second tutorial of the term. Two sections of the tutorial were used and were separated into two groups (1) the intervention group that followed the same procedure as the cohort II study, and (2) the control group for which their regular teaching assistant (TA) gave a lesson in acid–base chemistry instead of using the mechanisms module. The tutorial consisted of instruction in organic acid–base chemistry and determining the relative strength of acids and bases. The TA taught the students how to differentiate between weak and strong acids and bases and then the students were asked to complete some questions in small groups; the TA then went over the questions with the students. This specific tutorial and lesson were chosen because they are unrelated to the electron-pushing formalism and therefore served as a control group. Similar to the cohort II study, the participants were asked to provide informed consent to participate but any student was welcomed to attend without their data being used for research; Twenty-four and forty students provided consent in the intervention and control groups, respectively.

The pre-test and post-tests were identical to ensure that the tests were of equal difficulty (Figure 5.4). Questions 1–4 were aligned with LO1 (*Draw the arrows*); Questions 5–8 were aligned with LO2 (*Draw the products*). While all four LOs are believed to be important, due to time constraints, we focused on what we think are the LOs that are more essential to students' later success in analyzing mechanisms. During the workshop, students used the module and could ask questions of facilitators who were present. The intended LOs that would be tested were shared with the students and they were encouraged to focus on only those LOs during the session. Students worked individually, in pairs, or in small groups, according to their preference.

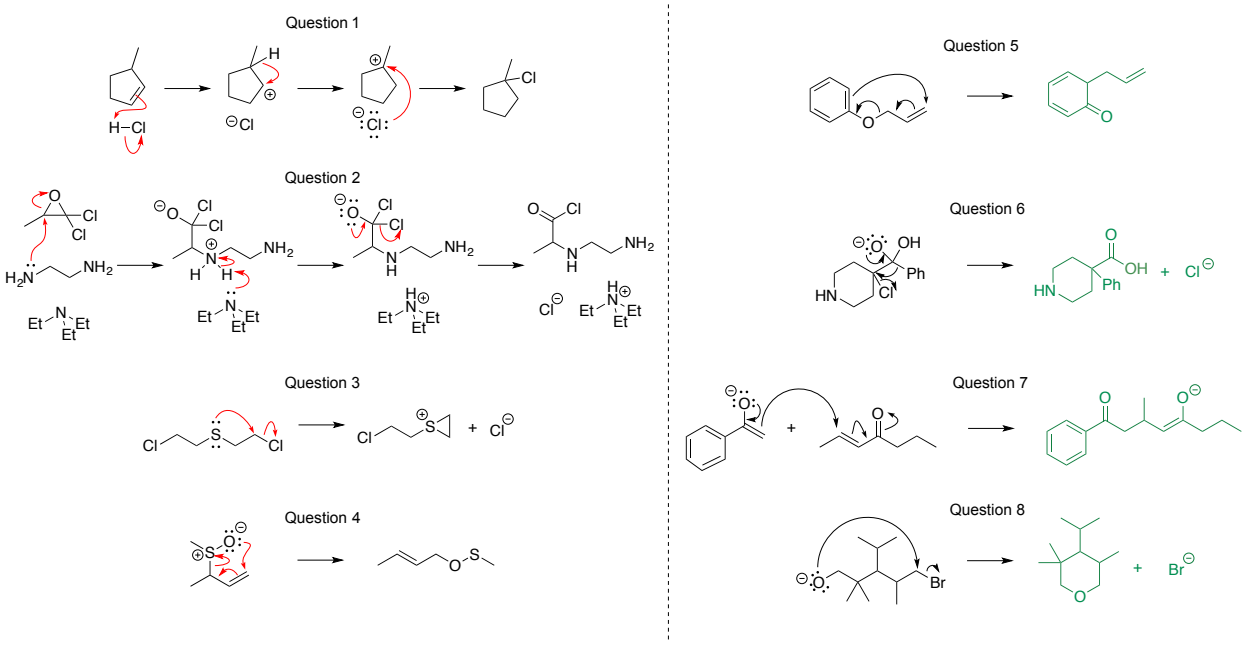



*Figure 5.4. Questions on the pre- and post-tests used in cohorts II and III, with Draw the Arrows questions (LO1) on the left with answers in red (bonds and electrons must be expanded if they are involved in the reaction step) and draw the Products questions (LO2) on the right with answers are in green.*

### 5.10.3 Data analysis

We analyzed the data according to the research questions, using R studio for statistical analyses. For RQ1, the worksheets were coded for the correctness of the answers. For the *Draw the arrows* questions, a point was given for each correct arrow, which are drawn in red in Figure 5.4. The *Draw the products* questions were coded per arrow. Interpreting an arrow correctly could be given up to two points, one for breaking a bond and one for making a bond, if relevant. In Question 5, for example, points would be associated with the long arrow if the answer showed the breaking the $\pi$ bond in the aromatic ring interpreted correctly and one point for forming the new C-C $\sigma$ bond correctly.

Some arrows, such as the small arrow in Question 8, where only worth one point since the arrow represents the breaking of a single (e.g., C-Br) bond and no new bonds formed. Similarly, the arrow for the collapse of the tetrahedral intermediate in Questions 6 and 7 was only worth one point because no bonds were broken, and one C-O $\pi$-bond formed.

For RQ2, we analyzed the following strategies for all questions: mapping, expanding the structure, redrawing the structure, and drawing non-bonding electrons. This analysis was done in the Cohort II study. We coded each strategy as being absent or present (even if it was not properly used). *Mapping* was used whenever a participant would mark (with a number, letter, shape, highlighter) part of a molecule to help situate it in the product. *Expanding the structure* was coded when implicit atoms and electrons/bonds were drawn. *Re-drawing the structure* implied that the student had re-drawn either the product or the starting material. *Drawing the non-bonding electrons* was coded when students drew the electrons on heteroatoms.

For RQ3, the errors on the tests were coded and analyzed for the Cohort II study (codebook is available Appendix 2.2). For the *Draw the arrows* LO*,* the following codes were used for each *Draw the arrows* question: correct, reversed, wrong, from atom/charge, vague, extra or did not attempt (Flynn and Featherstone, 2017). A *correct arrow* demonstrated the correct electron flow. A *reversed arrow* started from the electron deficient site and pointed toward the electron rich site. A *wrong arrow* started or ended at the wrong site on the molecule. An arrow coded *from atom/charge* demonstrated the correct electron flow from electron rich to electron poor but did not start from electrons; rather, the arrow started from a charge or an atom. *Missing arrow* was coded when a required arrow was absent, while *extra arrow* was coded if there were too many arrows present. Finally, if a student did not draw a single arrow, the *did not attempt* code was used.

For *Draw the products* questions, the errors were coded as in previous work (Flynn and Featherstone, 2017). For example, a *formal charge* error represented an instance with an extra, missing, or incorrect formal charge. A *placement error* was coded when the product was not correctly connected; a double bond or an atom was misplaced. *Transplanting electrons* represented a situation when electrons in a bond or on an atom were relocated to a new atom instead of bonding two atoms together.

### 5.10.4 Reliability of coding

Reliability was addressed via weekly debriefing session between the first and corresponding authors during which the codebook was developed and revised. Once the codebook was fully developed, one researcher coded the entire data set and second coder

analyzed 15% of the data to establish inter-rater reliability. The researchers obtained 91% agreement for the scores of the test with Krippendorff's $\alpha$ = 0.81, which is above the threshold of 0.80 (Krippendorff, 2004). For coding the errors, a 93% agreement was obtained, with a Krippendorff's $\alpha$ = 0.91. Similarly, for the strategies used, 89% agreement and Krippendorff's $\alpha$ = 0.90 was obtained. We were satisfied with the reliability of the data analysis, based on the aforementioned results.

## 5.11 Results and Discussion

### 5.11.1 Learning gains were observed across questions and LOs (RQ1)

In the Cohort II study, scores were significantly higher for the post-test than for the pre-test (Figure 5.5). Both question types had significant improvements. LO1's *Draw the arrows* scores were significantly higher on the post-test than on the pre-test. LO2's *Draw the products* scores were significantly higher on the post-test than on the pre-test. The highest gains in scores occurred with *Draw the products* questions.

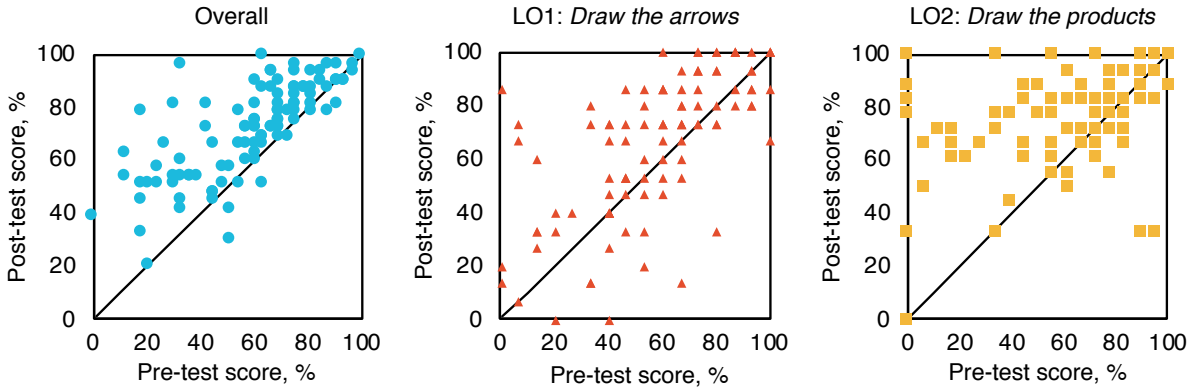



*Figure 5.5. Post-test versus pre-test scores: Overall (left, blue circles), for LO1: Draw the arrows (middle, red triangles), for LO2: Draw the products (right, yellow squares). N = 103. Cohort.*

Some of the students obtained a score of zero in some of the questions and their answers were examined further for validity. We found that these scores were all valid since they resulted from one of the following three scenarios: (1) some students focused on questions related to only one of the LOs, and thereby obtained a score of zero on the other LO; (2) the students answered the questions incorrectly; or (3) the students wrote a partial (and incorrect) answer on the questions. Since all of these students had answered at least one of the questions on the worksheet, their scores were included in the data analysis.

*Table 5.1. t-test statistics from comparing the pre-test and post-test scores for the overall score and scores on questions related to the individual LOs.*

| | Pre- Test | | Post-test | | | | |
|---|---|---|---|---|---|---|---|
| Independent variable | Mean (%) | SD (%) | Mean (%) | SD (%) | *t*(102) | p | Cohen's d |
| Overall score | 59.5 | 22.7 | 72.3 | 17.6 | 8.597 | < 0.001 | 0.729 |
| Scores from LO1: *Draw the arrows* | 58.3 | 26.1 | 66.7 | 26.9 | -3.992 | < 0.001 | 0.315 |
| Scores from LO2: Draw the products | 60.5 | 31.1 | 77.0 | 19.1 | -5.806 | < 0.001 | 0.863 |

Scores on Questions 4 and 8 had the largest improvements (Figure 5.6). Moreover, only 63% of participants attempted Question 4 and 57% attempted Question 8 on the pre-test, while 82% and 84% attempted those questions on the post-test, respectively.

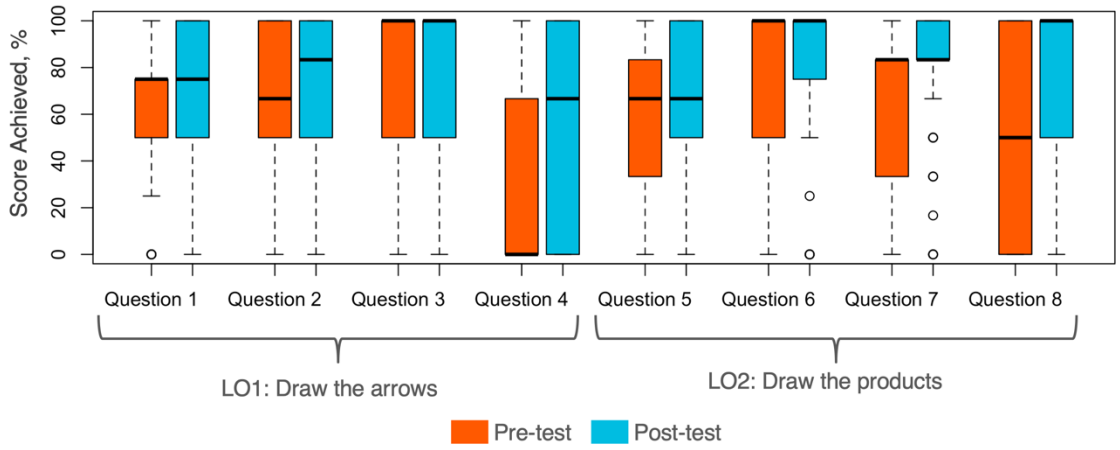



*Figure 5.6. Score distributions per question on the pre- and post-tests (N = 103). Cohort II.*

Normalized learning gains were also calculated and analyzed to account for the high scores of certain participants and a potential ceiling effect in Questions 3 and 6. Normalized learning gains account for the variance in the pre-test scores, the gain in score that is possible for each participant, and ceiling effects; they were calculated using Equation 1. The normalized learning gain calculation accounts for the fact that lower scores have more room for improvement and higher scores have less. A normalized learning gain of 1.0 indicates a perfect score on the post-test. A normalized learning gain of 0.0 indicates no improvement from the pre-test to the post-test, while a negative value shows deterioration of the score.

$$\text{Normalized learning gain} = \frac{(Posttest\ \%) - (Pretest\ \%)}{(100\%) - (Pretest\ \%)}$$

*Equation 1. Normalized learning gains.*

The normalized learning gains revealed improvements on the post-test from the pre-test as a whole (Figure 5.7). The biggest learning gains occurred in Question 8, which had a median normalized learning gain of 0.336. Questions 3 and 6 did not have high normalized learning gains since participants performed well on the pre-test, with means of 73% and 72%, respectively.

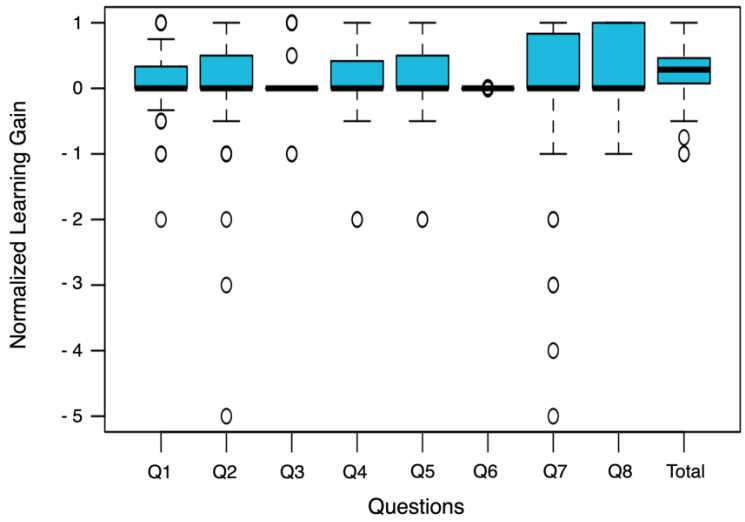



*Figure 5.7. Normalized learning gain scores, N = 103. Cohort II.*

### 5.11.2 Comparison between the intervention with non-EPF work

The Cohort III study was used to compare normalized learning gains between the intervention and control groups (Table 2, Figure 5.10). We analyzed the data using a Mann-Whitney test due to lack of normality in students' scores and learning gains in the control group. In the control group, there was no significant difference between the overall pre-test scores and the post-test scores. In contrast, the intervention group had significantly higher scores on the post-test than on the pre-test. The intervention group had LO2's *Draw the products* scores that were significantly higher on the post-test than on the pre-test. However, LO1's *Draw the arrows* scores were not statistically different between the pre-test and the post-test. These results we similar to the Cohort I study in which there were high learning gains in LO2 and the overall score and smaller learning gains in LO1.

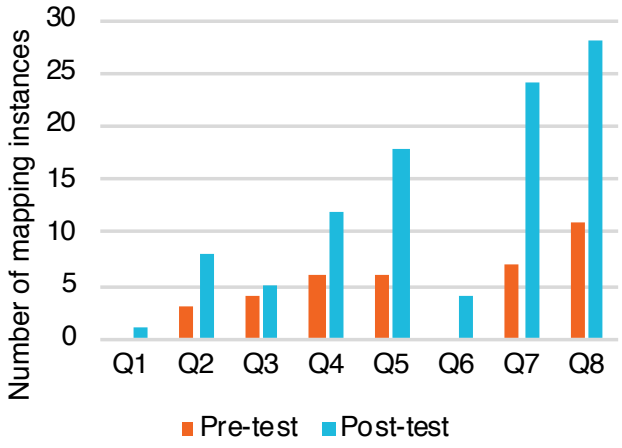



*Figure 5.8. Mapping instances found on the pre-test (orange) and post-test (blue) for each question. N = 103. Cohort II.*

*Table 5.2. Mann-Whitney test for comparison of pre-test scores and post-test scores for the intervention and the control groups.*

| Group | Independent variable | Pre-Test | | Post-Test | | U | p | r |
|---|---|---|---|---|---|---|---|---|
| | | Mean (%) | SD (%) | Mean (%) | SD (%) | | | |
| Intervention (N = 24) | Overall scores | 54.4 | 19.9 | 78.3 | 14.7 | 107.0 | < 0.001 | 0.548 |
| | Scores on LO1: Draw the arrows | 63.1 | 25.8 | 70.8 | 24.4 | 242.5 | 0.345 | - |
| | Scores on LO2: Draw the products | 47.2 | 33.8 | 84.5 | 11.8 | 79.5 | < 0.001 | 0.136 |
| Control (N = 40) | Overall scores | 60.0 | 24.0 | 62.5 | 20.3 | 776.0 | 0.817 | - |
| | Scores on LO1: Draw the arrows | 60.6 | 26.7 | 57.0 | 28.2 | 854.5 | 0.602 | - |
| | Scores on LO2: Draw the products | 59.4 | 29.0 | 67.1 | 22.6 | 719.0 | 0.434 | - |

*Table 5.3. Mann-Witney tests to compare the normalized learning gains from the control and the intervention groups.*

| Independent variable | Intervention (N = 24) | | Control (N = 40) | | U | p | r |
|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | | | |
| Overall normalized LG | - 0.076 | 0.391 | - 0.111 | 0.569 | 167.0 | < 0.001 | 0.548 |
| Normalized LG on LO1: *Draw the arrows* | - 0.250 | 0.254 | - 0.302 | 0.932 | 1102.0 | < 0.001 | 0.320 |
| Normalized LG on LO2: *Draw the products* | 0.525 | 0.450 | - 0.196 | 1.212 | 213.5 | < 0.001 | 0.462 |

We found a significant difference between the control and intervention group in the overall learning gains (Table 5.3), LO1 *Draw the arrows* normalized learning gains and LO2 *Draw the products* normalized learning gains. The large effect sizes calculated show that the intervention had a strong effect on students' learning gains compared to the control group. In

summary students in two different cohorts (cohort II and cohort III) had learning gains using the mechanism module over a short period of time.

### 5.11.3 Participants used common problem-solving strategies (RQ2)

To investigate RQ2, we analyzed the Cohort II study data for the strategies used. All of the strategies mentioned earlier (expanding the structure, re-drawing the structure, and mapping) are explained and used in the module. We anticipated seeing an increase in the use of those strategies from the pre-test to the post-test. We found that the only significant difference observed was in the use of mapping strategies. The other strategies (e.g., expanding structures, drawing non-bonding electrons) showed no significant change in use from the pre-test to the post-test (Figure 5.8). The lack of strategy use could be related to students not believing the strategies are useful, necessary, thinking that the strategies would take too much time to implement, or not having had enough time to see how they can be useful. In both question types, more students attempted the questions in the post-test than in the pre-test.

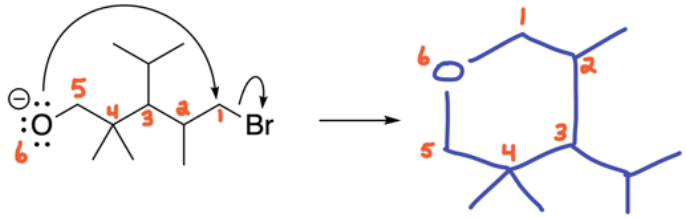



*Figure 5.9. Example of successful mapping used on the post-test, Question 8.*

Mapping was the strategy that was demonstrated significantly more often in the post-test (48%) than in the pre-test (20%), $\chi^2$ (1) = 23.361, $p < 0.001$, although still in relatively few questions overall (Figure 5.8). The module dedicated a lot of time to this subject and contained many mapping questions for students to practice. *Draw the products* questions (Q5 to Q8) had the biggest increase in use of mapping (example in Figure 5.9). Question 6 was done very well—average of 77% on the pre-test and 87% on the post-test—but very few students chose to map. We hypothesize that in those cases, participants chose not to map because they could visualize the answer or mapped in their heads. Question 8 had the largest increase in mapping (from 11 to 38 instances) as well as the largest learning gains (47% to 77%). Perhaps in this question, participants found value in mapping. A key decision point for students will be when to use such

strategies, much like how one needs to decide when heuristics are appropriate and when deeper analytical thinking is needed (Talanquer, 2014, 2017).

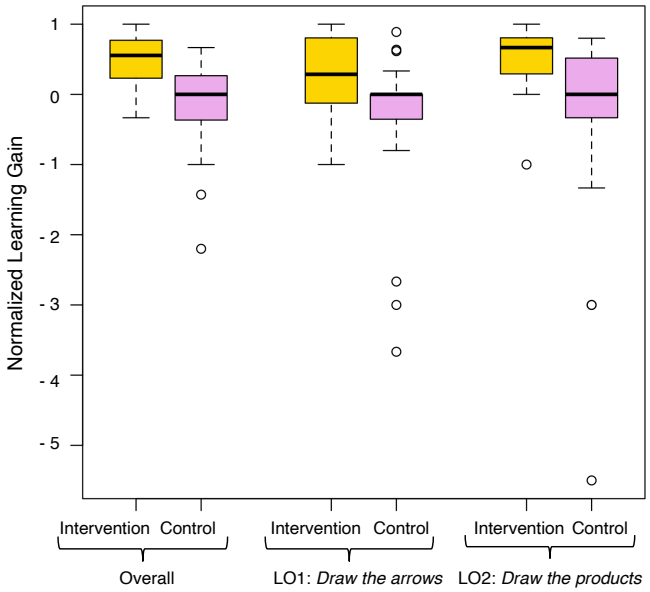



*Figure 5.10. Normalized learning gains for the intervention and control groups. Cohort II.*

Students who mapped in the pre-test (N = 18) had a higher mean score (70.1) than the participants who did not (N = 84, M = 57.0), $t(25) = 2.383$, $p = 0.034$. However, on the post-test both students who mapped (N = 49, M = 73.3) and did not map (N = 53, M = 71.5) obtained similar results, $t(100) = 0.513$, $p = 0.609$.

To analyze the effect of mapping on the students' learning gains, the participants were separated into four groups based on whether or not they mapped (Figure 5.11). Group 1 includes the sixteen participants who mapped on both the pre-test and the post-test. Group 2 includes the two participants who mapped on the pre-test and did not map on the post-test. Group 3 includes the 33 participants who did not map on the pre-test but mapped on the post-test. Group 4 includes the 51 participants who did not map on the pre-test or post-test.

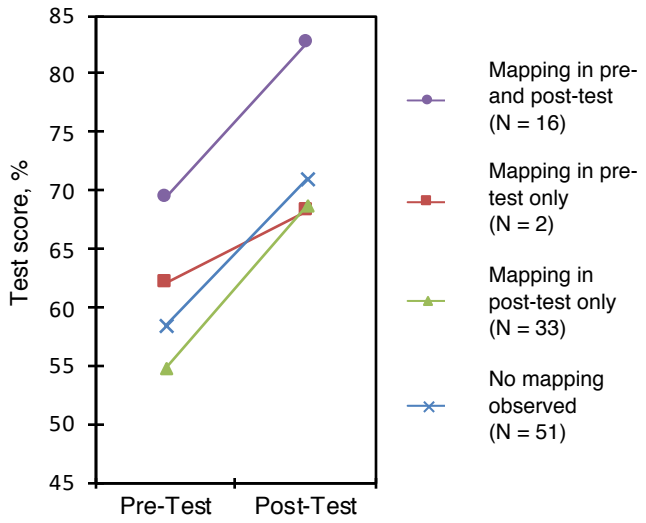



*Figure 5.11. Test scores, grouped according to whether mapping was observed in the pre-test and post-test. Cohort II.*

The participants in Group 3, who used mapping on the post-test but not on the pre-test showed learning gains of 13%, which is not statistically different than the average of 11% for all students.

A one-way ANOVA was done between Groups 1, 3, and 4 to determine the effect of the groups on normalized learning gains. Group 2 was excluded due to the low number of participants. There was no significant effect for the three groups [$F(2,97) = 0.941$, $p = 0.394$]. Participants who changed from not-mapping (pre-test) to mapping (post-test) had similar learning gains as the other groups. This analysis accounts for the overall score of the worksheet and not the specific questions related to mapping. Another limitation of the analysis is that the worksheets were not coded on whether the participants mapped properly or not, but simply whether they mapped at all. The increase in the use of the mapping strategy is nevertheless promising.

### 5.11.4 Common errors (RQ3)

To analyze the errors that occurred in the questions, we coded answers from Cohort II. For *Draw the arrows* questions (LO1), there was a significant increase in the number of correct arrows between the pre-test and post-test, $\chi^2$ (1) = 50.10, $p < 0.001$), consistent with the significant increase in learning gains (Figure 5.12). Overall, there was a similar percentage and type of errors on both tests, *i.e.*, no statistically significant difference (Figure 5.12). Another major

difference was that there were significantly more attempts on questions in the post-test than in the pre-test, $\chi^2(1) = 19.05$, $p < 0.001$.

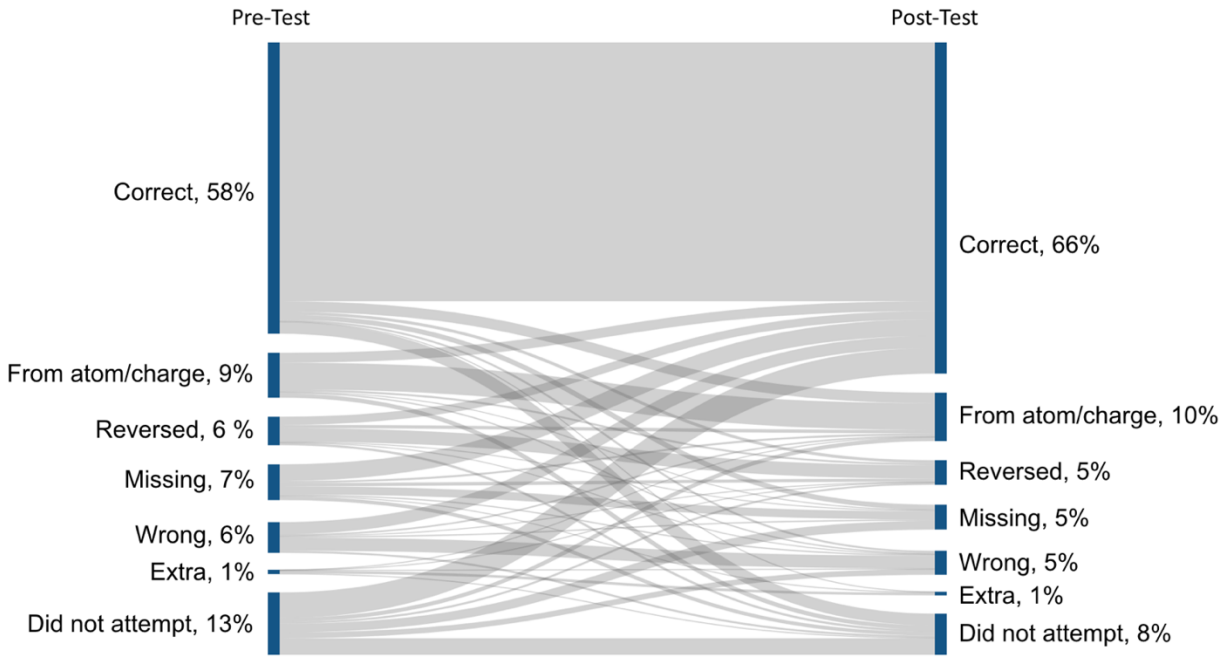



*Figure 5.12. For LO1, Draw the Arrows questions: Overview of correct answers, errors in arrows, and attempts between the pre-test (left) and the post-test (right). Each data point represents a single arrow, N = 1522.*

The most common type of error was drawing an arrow from an atom or charge (Figure 5.14). These arrows demonstrate the correct direction of electron movement; however, the start of the arrow did not begin with electrons and so we did not assess it to be correct. The module emphasizes that bonds are created by the movement of electrons—not atoms or charges; therefore, curved (EPF) arrows should start from electrons (Figure 5.13). In later years, we anticipate that students would adopt the common conventions used by organic chemists to draw EPF arrows from atoms or charges. Question 2 had the highest instance of this type of error, which had three of the six arrows start from electrons on a heteroatom, meaning that the students had to draw the electrons (*i.e.*, an extra step). In contrast, drawing arrows from atoms or charges was rarely observed in Question 1, which involved a hydride transfer.

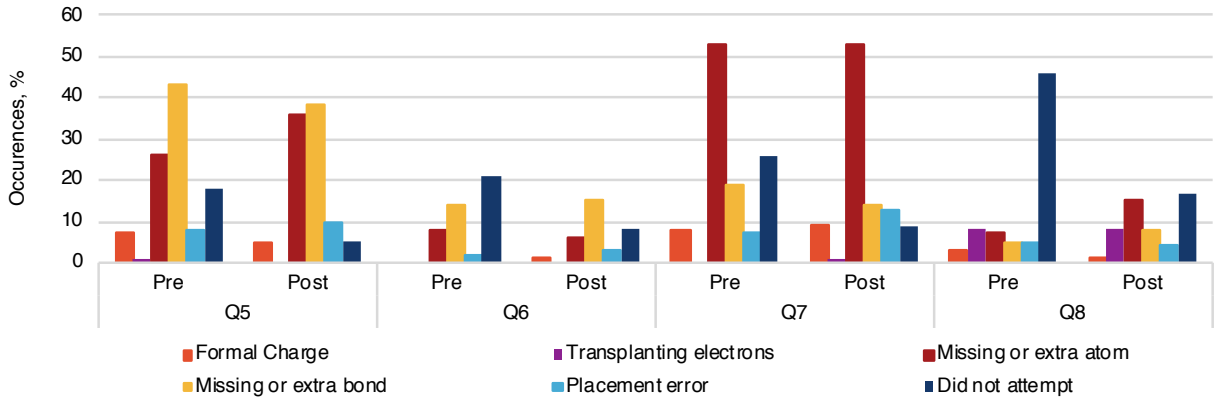



*Figure 5.13. Common errors found in answers for questions associated with LO2: Draw the products. N = 103. Cohort II.*

Questions associated with LO2 (*Draw the products*) had common errors consistent with previous work (Figure 5.13) (Flynn and Featherstone, 2017). Missing or extra bonds were prevalent in Question 5 (Figure 5.15). Questions tended to have specific types of errors; for example, answers to Question 7 were frequently missing a methyl group. Transplanting electrons was only observed in Question 8 (example in Appendix 2.2).

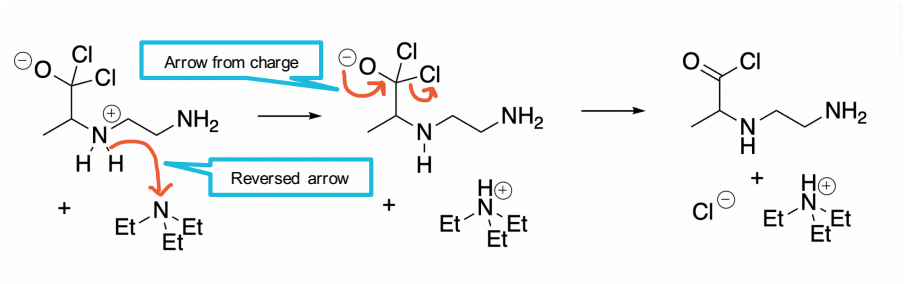



*Figure 5.14. Example of errors seen in Draw the arrows questions, specifically Question 2.*

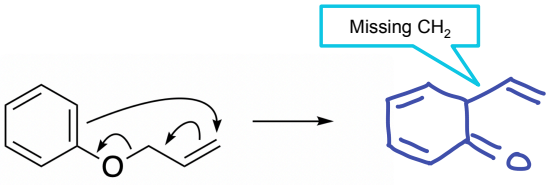



*Figure 5.15. Example of a missing atom error in a Draw the products question (LO2), specifically Question 5.*

For all the errors in the *Draw the products* questions, the "long" arrow seemed to be most challenging to interpret, while shorter arrows were interpreted correctly more often. For example, in Question 5, only 39% of students on the pre-test successfully made the bond from the long arrow, as opposed to 65% for the bottom left arrow and 62% for the bottom right arrow.

Comparing the arrow interpretations to each other showed significant difference between the long arrow and each of the shorter arrows, $\chi^2 = 23.59$, $p < 0.001$ and $\chi^2 = 22.88$, $p < 0.001$, respectively. There was no significant difference in interpretation between the two shorter arrows, $\chi^2 = 0.16$, $p = 0.344$. Similar results were found in the post-test, as well as Questions 7 and 8. These long arrows represent situations in which conformational changes are needed for the molecule to go from the conformation depicted in the starting material to the reactive conformation, which likely posed visualization and mental rotation challenges for students. Analogous difficulty was found between the same reactants shown in different conformations (Flynn and Featherstone, 2017). These results will need to be explored in more depth.

## 5.12 Conclusions

We found significant learning gains after students used the *Organic Mechanisms: Mastering the Arrows* module, an open education resource, in OrgChem101.com for one hour, focusing on the first two LOs of the module: *Draw the arrows*, given the starting materials and products of a reaction step (LO1) and *draw the products*, given the starting materials and electron-pushing arrows for that step (LO2), all for reactions the students had not previously seen. More students attempted the questions and used a mapping strategy in the post-test than in the pre-test. Students have reported a positive learning experience, belief that the module would help them in their organic chemistry courses, and greater ability to assess their skills related the LOs described herein (i.e., an aspect of metacognition) after using the module (Visser and Flynn, 2018). Students in the control group did not experience these learning gains. These findings are consistent with a previous educational evaluation of the *Nomenclature101* module in OrgChem101 (Flynn *et al.*, 2014; Bodé *et al.*, 2016). The strategies and errors observed in this study are also consistent with the types and relative prevalence found in exam analysis of electron-pushing formalism questions (Flynn and Featherstone, 2017).

### 5.12.1 Potential Limitations

This study did not make comparisons with other instructional methods so we cannot make claims about the effectiveness of this method compared to others, only that significant and large learning gains were observed in this study's context in a short time period. This study

demonstrated that this online tool is effective way for students to learn a key aspect of chemistry's language.

The learning gains of the cohort II study could be associated with time on task; analogous class time could have similar effects, as we found when evaluating the Nomenclature101 module in the same learning tool (Bodé *et al.*, 2016). Students' learning was measured over a short time period and so we do not make claims about the enduring nature of students' abilities or students' ability to transfer their skills to new situations. As with any learning, we expect that practice and use in context are essential for meaningful and enduring connections to made with other areas of chemistry. Students' learning was only studied with respect to the first two LOs of the module, although we hypothesize that the learning effects will be similar for intended LOs three (draw the transition state of a reaction step, given the starting materials and product of that step) and four (draw the mechanism of the reversed reaction, given the forward mechanism. Because the pre-test and post-test were identical, students may have remembered their answers from the pre-test. We think this limitation is unlikely since the pre-test was collected immediately after it was completed, and the questions were of sufficient difficulty and present in sufficient quantity to mitigate a memorization effect. Moreover, even if the students remembered their answers, a repeated individual response would tend to produce the same answer (*e.g.*, an incorrect answer would still be incorrect) if the intervention had no effect.

Another limitation of the study is that we did not collect data on which part of the module students finished or what they worked on during the workshop. The Classroom Observation Protocol for Undergraduate STEM courses (COPUS) was used to record what the students were generally doing during the workshop and showed that the students were working either by themselves or in group. However, we do not have data on what task exactly they were doing, and some students may have been working off-task or may not have completed the module.

### 5.12.2 Implications for teaching

This study demonstrated that the *OrgChem101.com: Mastering the arrows* module is a useful tool for students to learn and enhance their skills using the electron-pushing formalism. The mechanisms learning module could be used in a class setting, guiding students through its

use, and answering questions; alternatively, the expected LOs in the course could be provided to students and they could work on the module independently.

With either type of module use, summative assessment questions (*e.g.*, midterm and exam questions) should include aligned questions to demonstrate the value of the activities and module to the students as well as assess progress toward the intended LOs.

The module's approach and question types of address students' skill in interpreting the electron-pushing formalism, part of organic chemistry's language, but does not directly teach or assess concepts or reasons for reactivity observed. In uOttawa's curriculum, those concepts and reasons are addressed in other areas and question types, building off students' skills with the EPF (Flynn and Ogilvie, 2015; Raycroft and Flynn, 2017; Bodé and Flynn, 2018). The module could also be used in a traditional curriculum to help students learn EPF skills and allow the students to practice on new questions focused solely on the EPF. The learning module also teaches strategies that are correlated with greater success for organic synthesis problems (Bodé and Flynn, 2016). The increase in the usage of mapping highlights the usefulness of the learning module in modelling these strategies for students.

### 5.12.3 Implications for research

The module is an ever-improving learning tool, therefore further research into its usefulness would help enhance the module. Further research into why students use strategies and interviews about the thought-process students engage in while solving these problems could help the development team enhance the module. The design and development of a module such as this is always iterative, and the data obtained will guide the next iteration to improve the module further.

The effect of this module could also be studied for students currently in a traditional curriculum. Future research could investigate the effects of gaining fluency in organic chemistry's symbolism on students' abilities to learn new concepts in organic chemistry that require using organic chemistry's symbolism. Future research could also investigate the existence and extent of benefits for a variety of learners (e.g., gender differences, effects of technological proficiency on learning).

## 5.13 References


Anzovino M. E. and Lowery Bretz S., (2015), Organic Chemistry Students' Ideas About Nucleophiles and Electrophiles: The Role of Charges and Mechanisms. *Chemistry Education Research and Practice*, **16**(4), 797–810.

Ausubel D. P., Novak J. D., and Hanesian H., (1978), *Educational psychology: A cognitive view*, Rinehart and Winston (eds.) Holt.

Becker N., Noyes K., and Cooper M. M., (2016), Characterizing Students' Mechanistic Reasoning about London Dispersion Forces. *Journal of Chemical Education*, **93**(10), 1713–1724.

Bhattacharyya G., (2013), From Source to Sink: Mechanistic Reasoning using the Electron-Pushing Formalism. *Journal of Chemical Education*, **90**(10), 1282–1289.

Bhattacharyya G., (2014), Trials and Tribulations: Student Approaches and Difficulties with Proposing Mechanisms using the Electron-Pushing Formalism. *Chemistry Education Research and Practice*, **15**(4), 594–609.

Bhattacharyya G. and Bodner G. M., (2005), "It Gets Me to the Product": How Students Propose Organic Mechanisms. *Journal of Chemical Education*, **82**(9), 1402.

Bodé N. E., Caron J., and Flynn A. B., (2016), Evaluating Students' Learning Gains and Experiences from Using Nomenclature101.com. *Chemistry Education Research and Practice*, **17**(4), 1156–1173.

Bodé N. E., Deng J. D., and Flynn A. B., (2019), Getting Past the Rules and to the WHY: Causal Mechanistic Arguments When Judging the Plausibility of Organic Reaction Mechanisms. *Journal of Chemical Education*, **96**(6), 1068-1082.

Bodé N. E. and Flynn A. B., (2016), Strategies of Successful Synthesis Solutions: Mapping, Mechanisms, and More. *Journal of Chemical Education*, **93**(4), 593–604.

Bretz S. L., (2001), Novak's Theory of Education: Human Constructivism and Meaningful Learning. *Journal of Chemical Education*, **78**(8), 1107.

Brown P. C., Roediger H. L., and McDaniel M. A., (2014), *Make it stick: the science of successful learning*, Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

Caspari I., Weinrich M., Sevian H., and Graulich N., (2017), This Mechanistic Step is "Productive": Organic Chemistry Students' Backward-Oriented Reasoning. *Chemistry Education Research and Practice*, **19**(1), 42–59.

Cheng M. and Gilbert J. K., (2009), Towards a Better Utilization of Diagrams in Research into the Use of Representative Levels in Chemical Education, in *Multiple Representations in Chemical Education*, Springer, pp. 55–73.

Cooper M. M., Kouyoumdjian H., and Underwood S. M., (2016), Investigating Students' Reasoning about Acid–Base Reactions. *Journal of Chemical Education*, **93**(10), 1703–1712.

Council N. R., (2000), *How People Learn: Brain, Mind, Experience, and School: Expanded Edition*, The National Academies Press.

Craig A. F., Koch D. L., Buffington A., and Grove N., (2012), Narrowing the Gap? Revisiting Publication Rates in Chemistry Education. *Journal of Chemical Education*, **89**(12), 1606–1608.

DeCocq V. and Bhattacharyya G., (2019), TMI (Too Much Information)! Effects of Given Information on Organic Chemistry Students' Approaches to Solving Mechanism Tasks. *Chemistry Education Research and Practice*, **20**(1), 213–228.

DeFever R. S., Bruce H., and Bhattacharyya G., (2015), Mental Rolodexing: Senior Chemistry Majors' Understanding of Chemical and Physical Properties. *Journal of Chemical Education*, **92**(3), 415–426.

Ferguson R. and Bodner G. M., (2008), Making Sense of the Arrow-Pushing Formalism Among Chemistry Majors Enrolled in Organic Chemistry. *Chemistry Education Research and Practice*, **9**(2), 102–113.

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. *Journal of Chemical Education*, **91**(11), 1855–1859.

Flynn A. B., Caron J., Laroche J., Richard G., Bélanger M., and Featherstone R., (2016), Orgchem101.com: An Organic Chemistry and Metacognitive Skill and Concept Building Tool. OrgChem101 Learning Lab. http: //orgchem101.com/

Flynn A. B. and Featherstone R. B., (2017), Language of Mechanisms: Exam Analysis Reveals Students' Strengths, Strategies, and Errors when using the Electron-Pushing Formalism (Curved Arrows) in New Reactions. *Chemistry Education Research and Practice*, **18**(1), 64–77.

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. *Journal of Chemical Education*, **92**(5), 803–810.

Flynn A. B. and Visser R., (2018), Developing Open Educational Resources in French and English for Students of Organic Chemistry at the University of Ottawa, Canada. *Contact North, Contact Nord*.

Galloway K. R. and Bretz S. L., (2015), Measuring Meaningful Learning in the Undergraduate General Chemistry and Organic Chemistry Laboratories: A Longitudinal Study. *Journal of Chemical Education*, **92**(12), 2019–2030.

Galloway K. R., Leung M. W., and Flynn A. B., (2019), Patterns of Reactions: A Card Task to Investigate Students' Organization of Organic Chemistry Reactions. *Chemistry Education Research and Practice*, **20**(1), 30–52.

Galloway K. R., Stoyanovich C., and Flynn Alison B, (2017), Students' Interpretations of Mechanistic Language in Organic Chemistry Before Learning Reactions. *Chemistry Education Research and Practice*, **18**(2), 353–374.

Gilbert J. K., (2008), *Visualization: An Emergent Field of Practice and Enquiry in Science Education*, Gilbert J. K., Reiner M., and Nakleh M. (eds.).

Gilbert J. K. and Treagust D. F., (2009), *Multiple Representations in Chemical Education*, Gilbert J. K., and Treagust D. F. (eds.) Springer.

Graulich N., (2015), The Tip of the Iceberg in Organic Chemistry Classes: How do Students Deal with the Invisible? *Chemistry Education Research and Practice*, **16**(1), 9–21.

Grove N. P., Cooper M. M., and Cox E. L., (2012), Does Mechanistic Thinking Improve Student Success in Organic Chemistry? *Journal of Chemical Education*, **89**(7), 850–853.

Grove N. P., Cooper M. M., and Rush K. M., (2012), Decorating with Arrows: Toward the Development of Representational Competence in Organic Chemistry. *Journal of Chemical Education*, **89**(7), 844–849.

Grove N. P. and Lowery Bretz S., (2012), A Continuum of Learning: From Rote Memorization to Meaningful Learning in Organic Chemistry. *Chemistry Education Research and Practice*, **13**(3), 201–208.

Hannafin M. J., (2012), Student-Centered Learning, in *Encyclopedia of the Sciences of Learning*, Seel N. M. (ed.)., Springer US, pp. 3211–3214.

Johnstone A. H., (1982), Macro- and Micro-Chemistry. *School Science Review*, **64**, 377–379.

Johnstone A. H., (2000), Teaching of Chemistry - Logical or Psychological? *Chemistry Education: Research and Practice*, **1**(1), 9–15.

Kalyuga S., Ayres P., Chandler P., and Sweller J., (2003), Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist*, **38**, 23–31.

Kozma R. B. and Russell J., (1997), Multimedia and Understanding: Expert and Novice Responses to Different Representations of Chemical Phenomena. *Journal of Research in Science Teaching*, **34**(9), 949–968.

Kraft A., Strickland A. M., and Bhattacharyya G., (2010), Reasonable Reasoning: Multi-Variate Problem-Solving in Organic Chemistry. *Chemistry Education Research and Practice*, **11**(4), 281–292.

Krippendorff K., (2004), Reliability in Content Analysis. *Human Communication Research*, **30**(3), 411–433.

Lapierre K. R., Streja N., and Flynn A. B., (2022), Investigating the Role of Multiple Categorization Tasks in a Curriculum Designed around Mechanistic Patterns and Principles. *Chemistry Education Research and Practice*.

Lieber L. and Graulich N., (2020), Thinking in Alternatives—A Task Design for Challenging Students' Problem-Solving Approaches in Organic Chemistry. *Journal of Chemical Education*, **97**(10), 3731–3738.

Mayer R. E., Heiser J., and Lonn S., (2001), Cognitive Constraints on Multimedia Learning: When Presenting More Material Results in Less Understanding. *Journal of educational psychology*, **93**(1), 187

Mayer R. E. and Moreno R., (2003), Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist*, **1520**(38), 43–52.

Moreira P., Marzabal A., and Talanquer V., (2018), Using a Mechanistic Framework to Characterise Chemistry Students' Reasoning in Written Explanations. *Chemistry Education Research and Practice*, Advance Article.

National Research Council, (2012), A Framework for K-12 Science Education.

National Research Council (NRC), (2012), *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*, Singer S. R., Nielsen N. R., and Schweingruber H. A. (eds.) The National Academies Press.

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

Raycroft M. and Flynn A. B., (2017), Next Steps Toward Improving the Organic Chemistry Curriculum at the University of Ottawa, in *Canadian Society for Chemistry: 100th Canadian Chemistry Conference and Exhibition*.

Rushton G. T., Hardy R. C., Gwaltney K. P., and Lewis S. E., (2008), Alternative Conceptions of Organic Chemistry Topics Among Fourth Year Chemistry Students. *Chemistry Education Research and Practice*, **9**(2), 122–130.

Russ R. S., Scherr R. E., Hammer D., and Mikeska J., (2008), Recognizing Mechanistic Reasoning in Student Scientific Inquiry: A Framework for Discourse Analysis Developed from Philosophy of Science. *Science Education*, **92**(3), 499–525.

Schunk D., (2016), *Learning Theories: An Educational Perspective*, 6th ed. Pearson.

Stoyanovich C., Gandhi A., and Flynn A. B., (2015), Acid − Base Learning Outcomes for Students in an Introductory Organic Chemistry Course. *Journal of Chemical Education*, **92**(2), 220–229.

Sweller J., (1999), *Instructional design in technical areas*, Australia: ACER Press.

Taber K. S., (2009), Learning at the Symbolic Level, in *Multiple Representations in Chemical Education*, Springers Netherlands, pp. 75–105.

Talanquer V., (2014), Chemistry Education: Ten Heuristics to Tame. *Journal of Chemical Education*, **91**(8), 1091–1097.

Talanquer V., (2017), Concept Inventories: Predicting the Wrong Answer May Boost Performance. *Journal of Chemical Education*, **94**(12), 1805–1810.

Talanquer V., (2011), Macro, Submicro, and Symbolic: The Many Faces of the Chemistry "Triplet." *International Journal of Science Education*, **33**(2), 179–195.

Taskin V. and Bernholt S., (2014), Students' Understanding of Chemical Formulae: A review of empirical research. *International Journal of Science Education*, **36**(1), 157–185.

Vasilyeva N., Blanchard T., and Lombrozo T., (2018), Stable Causal Relationships Are Better Causal Relationships. *Cognitive Science*, **42**(4), 1265–1296.

Visser R. and Flynn A. B., (2018), What are Students' Learning and Experiences in an Online Learning Tool Designed for Cognitive and Metacognitive Skill Development? *Collected Essays on Learning and Teaching*, **11**, 129–140.

Webber D. M. and Flynn A. B., (2018), How Are Students Solving Familiar and Unfamiliar Organic Chemistry Mechanism Questions in a New Curriculum? *Journal of Chemical Education*, **95**(9), 1451–1467.

Weinrich M. L. and Sevian H., (2017), Capturing Students' Abstraction while Solving Organic Reaction Mechanism Problems Across a Semester. *Chemistry Education Research and Practice*, **18**(1), 169–190.

Weinrich M. and Talanquer V., (2015), Mapping Students' Conceptual Modes when Thinking about Chemical Reactions used to Make a Desired Product. *Chemistry Education Research and Practice*, **16**, 561–577.

# Chapter 6. Learning outcomes for Delocalization: How are they Intended, Enacted, and Achieved?

*"Students also commonly emphasized an operational rather than conceptual definition of resonance in their written explanations. Instead of discussing fundamental features of the concept, such as when resonance can occur and the fact that the atomic composition and connectivity remains unchanged between resonance contributors, students focused on how electrons are moved when drawing resonance structures. […] When describing the influence of resonance on reactivity, students seldom provided specific applications or contexts for considering reactivity". (Brandfonbrener et al., 2021)*

## 6.1 Delocalization: an important but confusing concept

In this chapter, I explore how the concept of electron delocalization is taught and achieved in Organic Chemistry courses. Electron delocalization is a critical concept that explains molecular properties and reactivity in many contexts, including materials chemistry, pharmaceutical science, solar cells, and biochemistry. The underlying concepts of resonance also explain the reactivity of functional groups and many reactions can be explained using electron delocalization concepts, including electrophilic aromatic substitution, Diels-Alder, and conjugate addition reactions. As such delocalization is a foundational aspect of organic chemistry.

Delocalization is sometimes called resonance by chemists; however, that can be seen as a misleading word (Kerber, 2006) and there has been debate over the use of the term (Jensen, 2006; Kerber, 2006; Truhlar, 2007). The IUPAC (2014b) defines "resonance" as: "the term refers to the representation of the electronic structure of a molecular entity in terms of contributing structures…" and therefore includes any Lewis structure that contributes to the wavefunctions, even unlikely structures. This definition would mean that delocalization across sigma bonds are resonance structures and that hyperconjugation would also fall under resonance (Mullins, 2012).

Therefore, resonance structures do not solely differ on placements of π orbitals electrons and non-bonding electrons but are related to any structure that contributes to the wavefunction. The term delocalization most accurately described electrons being shared over multiple atoms in π orbital delocalization (IUPAC, 2014a).

Delocalization of electron occurs when electrons are not localized to a single atom but instead are spread through a system. For delocalization to occur, a molecule must have molecular orbital formed from several adjacent *p* atomic orbitals. The formation of the molecular orbital lowers the energy of the bonding orbitals, making the overall molecule more stable.

Delocalization is typically represented as several different Lewis structures with location of non-bonding electrons, but the discrete resonance structures do not exist (Figure 6.1, Left). A molecule is a combination of all the resonance structures, which can be represented using the resonance hybrid (Figure 6.1, Right); however, the hybrid structure does not show the number of electrons that are delocalized.

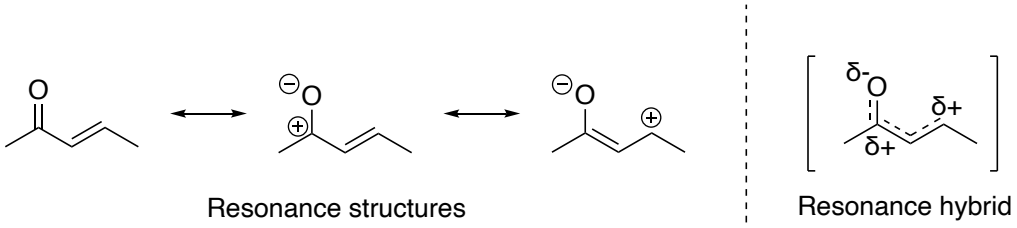



*Figure 6.1. Resonance structures of an enone (left) and its resonance hybrid (right)*

For the remainder of this chapter, I will be using delocalization to refer to electron delocalization. However, I will use the term resonance when quoting other work or participants. I will also use the terms "resonance structures" and "resonance hybrid" when referring to those representations.

## 6.2 Context: It is unclear what students should know about delocalization

While educators have reported that delocalization as a fundamental organic chemistry concept, and one of the hardest concepts for students to understand (Duis, 2011), learners often struggle with understanding and using the concept (Betancourt-Pérez *et al.*, 2010; Kim *et al.*, 2019; Xue and Stains, 2020; Brandfonbrener *et al.*, 2021). Multiple ways of teaching resonance have been proposed (Gero, 1954; Delvigne, 1989; Starkey, 1995; Silverstein, 1999; Lin, 2007; Kim

*et al.*, 2019), along with some course expectations (Betancourt-Pérez *et al.*, 2010). Four LOs have been proposed: (1) Draw the curved arrows for electron movement between resonance structures, (2) Draw the alternating Lewis structures, (3) Identify the most stable Lewis structure, and (4) Draw the resonance hybrid (Betancourt-Pérez *et al.*, 2010). Those LOs only focus on the initial teaching of the subject but lacks LOs related to the using the concept. Similarly, there has been some reports of how some learners conceptualize resonance structures has alternating or in equilibrium (Taber, 2002; Kim *et al.*, 2019; Xue and Stains, 2020), but few reports on how learner skills using delocalization in context. My goal is to investigate what students should know about delocalization and what skills learners have using delocalization.

## 6.3 Students' conceptualization and skills using delocalization

### 6.3.1 Learners may show alternate conceptions of what delocalization is

Previous investigation of students' understanding of delocalization concepts found that students struggled to conceptualize delocalization (Taber, 2002). Participants often thought of benzene's resonance structures as alternating between two states or as an electron reservoir *inside* the ring (Taber, 2002). Another study reported that students have multiple alternate conceptions about resonance and struggle to understand what the representation portrays, equating resonance with a chemical equilibrium (Xue and Stains, 2020). Rather than understanding that resonance contributors represent a unified structure, participants in one study from a general chemistry course viewed resonance structures as viewing a molecule from a different perspective or that a change had occurred between the structures (Kim *et al.*, 2019). However, after an intervention in which students drew their own representations of resonance and determined the pros and cons of the representations, there was a significant shift in students' view of resonance structures toward a unitization view (average of represented structures).

### 6.3.2 Representational competency

Learners may have difficulty using delocalization because of the representations used, and the skills required to understand those representation. Chemistry can be difficult for learners because of the different types of representations use, for which learners may not have the skills to use (Kozma and Russell, 1997; George M Bodner and Domin, 2000; Melanie M Cooper *et al.*,

2010; Sim and Daniel, 2014). Representational competency is "set of skills and practices that allow a person to reflectively use a variety of representations or visualizations, singly and together, to think about, communicate, and act on chemical phenomena in terms of underlying, perceptual physical entities and processes"(Kozma and Russell, 1997). The skills associated with representational competency are: (1) using representations to describe a chemical phenomenon in terms of molecules and molecular processes, (2) selecting an appropriate representation and explain why it is appropriate, (3) interpreting features of a representation, (4) making connections across different related representations, (5) understanding that representations does not always show the entirety of a concept, (6) describing limitations and affordances of different representations, and (7) using representations to draw inferences and make predictions about chemical phenomena (Kozma and Russell, 2005).

From previous finding related to delocalization, we can hypothesize that some of the alternate conceptions about delocalization stem from the representations used. Specifically, in terms of delocalization learners have been reported to struggle with interpreting the representation (Taber, 2002; Kim *et al.*, 2019; Xue and Stains, 2020), connecting related representations (i.e., the resonance structures and the resonance hybrid) (Betancourt-Pérez *et al.*, 2010; Xue and Stains, 2020), and understanding that a representation has limitation and affordance (Xue and Stains, 2020). For example, when asked to draw a resonance hybrid from resonance structures, 3% of participant drew the resonance hybrid correctly while 14% drew the correct bonds but missed partial charges, and more than half (53%) drew a major or minor contributor (Xue and Stains, 2020).

### 6.3.3 Learners' skills using delocalization concepts

There are several skills, both representational and conceptual, with using delocalization – from identifying when to use it, drawing structures or resonance hybrid, to using it to dictate reactivity/properties. Students may not even think to use delocalization unless when prompted to do so (Bodner and Domin, 2000; Finkenstaedt-Quinn *et al.*, 2020; Petterson *et al.*, 2020; Watts *et al.*, 2020). Students may not be aware of when delocalization is required, or that they chose not to use it. Several students had to be prompted to realize that delocalization could occur (Finkenstaedt-Quinn *et al.*, 2020). Once participants knew to use delocalization, they were mostly

successful in applying the rules to determine the radiochemistry of an addition reaction (Finkenstaedt-Quinn *et al.*, 2020).

When determining the stability of carbocations, participants seem to realize that resonance plays an important factor when determining stability or acidity (Finkenstaedt-Quinn *et al.*, 2020; Petterson *et al.*, 2020). When deprotonating a dicarbonyl, students used delocalization to determine the site of deprotonation, however when looking at imidazole students knew to use delocalization to determine the site but used the concept on the reactants as opposed to the product (Petterson *et al.*, 2020). The results point to the idea that the participants understood that delocalization is an important factor when determining stability or acidity/basicity but may not have applied it properly to certain problems (Petterson *et al.*, 2020).

## 6.4 Learning tool: LOs to obtain constructive alignment

Learning outcomes (LOs) are the knowledge, skills, and values that students *demonstrate* following a learning experience such as a module, section, course, program, or degree (Collis and Biggs, 1986; Brabrand and Dahl, 2009). LOs can be thought of as intended, enacted, and achieved. The *intended* aspect is the knowledge, skills, and values students are expected to be able demonstrate at the end of the course. The *enacted* aspect is how the subject matter is presented to students; it is separated into three component: taught, practiced, and assessed (Carle and Flynn, 2020; Raycroft and Flynn, 2020). The *achieved* aspect is how students demonstrate mastery of the subject matter (i.e., on assessments).

Ideally, the learning environment is inclusively designed in a way that aligns the intended LOs, with instruction and assessment, known as constructive alignment (Biggs and Tang, 2011; Kandlbinder, 2014). In doing so, the focus shifts from what the educator does to supporting the students' learning needs. Educators can design activities, assessments, and a learning environment to support students' argumentation skills through backward design (Wiggins and McTighe, 1998) by: (1) identifying the intended LOs, (2) determining practice opportunities and acceptable evidence of achievement, and (3) planning the learning experience. LOs can be used to clearly communicate expectations and as metacognitive tools for students (O'Connor et al., 2021). They can be communicated in multiple ways: in a course syllabus or at the beginning/end of each lecture, class, *etc*. Once the LOs are established, we can analyze how they are enacted,

including instruction, the environment, students' motivations, and how the LOs are taught, practiced, and assessed.

## 6.5 Goals and research questions

The goal of this research is to (1) identify the delocalization LOs, (2) analyze how they were enacted, and (3) analyze how they were achieved. This research was guided by the following research questions:

- **RQ1:** What are the essential intended LOs related to electron delocalization, by the end of a two-semester sequence in organic chemistry?
- **RQ2:** How are the proposed delocalization LOs being enacted?
    - RQ2a: How are the delocalization LOs being taught in introductory organic chemistry textbooks?
    - RQ2b: How are the delocalization LOs being practiced in introductory organic chemistry textbooks?
    - RQ2c: How are the delocalization LOs being assessed in organic chemistry courses?
- **RQ3:** How are the delocalization LOs being achieved by students on summative assessments?
    - RQ3a: To what extent are students achieving the LOs
    - RQ3b: What strategies are students using when answering delocalization related exam questions?
    - RQ3c: What errors are students making when answering delocalization related exam questions?

The remainder of this chapter consists of two manuscripts, the first of which addresses RQ1 and RQ2, while the second address RQ3.

## 6.6 Published contribution: What are the delocalization learning outcomes, and how are they enacted

*Essential learning outcomes for delocalization (resonance) concepts: How are they taught, practiced, and assessed in organic chemistry?*

Myriam S. Carle, and Alison B. Flynn

**Abstract:** The concept of delocalization (*i.e.*, resonance) is fundamental concept in organic chemistry but essential learning outcomes (LOs) have not previously been proposed nor has there been an analysis of how resonance is taught, despite indications in the literature that students have many non-canonical ideas about the concepts. To address this deficit, we first developed a set of ten LOs believed to be essential to the concept of delocalization in organic chemistry, especially for students' later success. Next, we analyzed how these LOs are being taught, practiced, and assessed in common textbooks and in a sample of exams. Five themes emerged from the analysis: (1) Several of the essential intended LOs we identified are not represented in the textbooks' teaching explanations, practice questions, or professors' assessments; (2) The concepts related to delocalization are often taught, practiced, and assessed without associated justifications; (3) There is a large gap between when delocalization is taught and when it is used in context; (4) The link between delocalization and other concepts (*e.g.*, reactivity) is not explicitly explained in most teaching materials; and (5) The language used around delocalization may be misleading (*e.g.* resonance, stability). Our analysis identified areas in which delocalization education could be improved, including with respect to teaching, practice opportunities, and assessing the concepts.

**Reproduced with permission from**: Carle M. S. and Flynn A. B., (2020), Essential learning outcomes for delocalization (resonance) concepts: How are they taught, practiced, and assessed in organic chemistry? *Chemistry Education Research and Practice,* 21(2), 622–637. http://doi.org/10.1039/C9RP00203K

**Author contribution:** MSC and ABF brainstorm the idea and design of the study. MSC did the data collection, data analysis, drafted the manuscript, and edited the manuscript based on

feedback. The authors discussed all phases of the work. ABF provided guidance and feedback throughout the entirety of the project.

## 6.7 Introduction

### 6.7.1 Problem and research goals

Resonance is a critical concept that explains molecular properties and reactivity in many contexts, including materials chemistry, pharmaceutical science, solar cells, and biochemistry. Conjugated polymers have unique properties due to the delocalization of electrons and these polymers can harness energy from light and be used in solar cells (Günes *et al.*, 2007; Jiang and McNeill, 2017). In the pharmaceutical industry, every top selling drug in 2016 contained a moiety with electron delocalization, except salts (Smith *et al.*, 2016). The underlying concepts of resonance also help explain the reactivity of functional groups. The reactivity in many reactions can be explained using electron delocalization concepts, including electrophilic aromatic substitution, Diels-Alder, and conjugate addition reactions. In other words, resonance is a foundational aspect of organic chemistry.

Educators have reported that resonance is a fundamental organic chemistry concept and one of the hardest concepts for students to understand (Duis, 2011) and ways of teaching resonance have been proposed (Gero, 1954; Delvigne, 1989; Starkey, 1995; Silverstein, 1999; Lin, 2007). However, these proposed ways of teaching delocalization did not outline what students should learn (*i.e.*, the expected LOs), what the students struggle to conceptualized, or investigated the impacts for student learning. More chemistry education research into domain-specific concepts is required, such as resonance, aromaticity and hyperconjugation (Graulich, 2015).

In the present study, we analyzed the teaching and learning of delocalization concepts by proposing essential LOs and analysing how concepts and skills addressed in those LOs are commonly being taught, practiced, and assessed.

### 6.7.2 The few existing studies show students' challenges conceptualizing resonance

One previous investigation of students' understanding of resonance concepts found that students struggled to conceptualize resonance (Taber, 2002). That study focused on how participants conceptualized molecular orbitals and electron delocalization and found that students often thought of benzene's resonance structures as alternating between two states (Figure 6.2, left). The idea of "alternating" structures can be due to the everyday meaning of the

word resonance, which means oscillating or alternating. In chemistry, resonance does not involve molecules alternating or oscillating between forms; their structure is a combination or hybrid of resonance contributors. To represent that hybrid, a common representation of aromatics involves a circle in the middle that represents the delocalized electrons. Taber found that the second representation type was also problematic, seen by many students as an electron reservoir *inside* the ring (Figure 6.2, right).

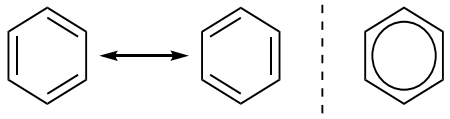


*Figure 6.2. Benzene can be represented as two resonance structure (left) or with a circle to represent aromaticity (right)*

Betancourt-Perez analyzed students' answers on four resonance related tasks: (1) Draw the curved arrows for electron movement between resonance structures, (2) Draw the alternating Lewis structures, (3) Identify the most stable Lewis structure, and (4) Draw the resonance hybrid (Betancourt-Pérez *et al.*, 2010). The authors identified common mistakes for each of the tasks and found a significant correlation between tasks 1 to 3 with the students' exam scores in their first two semesters of Organic Chemistry. The authors also used a multiple-regression analysis to determine that proficiency in task 1 was the best predictor for student exam scores and that mastery of tasks 1,2 and 3 were a good predictor for course grades.

Rather than understanding that resonance contributors represent a unified structure, student participants from a general chemistry course viewed resonance structures as dimensionalized (viewing a molecule from a different perspective) or segmented (change had occurred between the structures)(Kim *et al.*, 2019). However, after an intervention in which students drew their own representations of resonance and determined the representations' pros and cons, there was a significant shift in students' view of resonance structures toward a unitization view (average of represented structures).

To more precisely communicate the idea that electrons are delocalized and not alternating between structures, we use the term "delocalization" herein except when talking about resonance structures or the resonance hybrid. While the term "mesoisomeric structures"

could be used, as opposed to “resonance structures”, we found that the term is so uncommon (*e.g.*, none of the textbook analyzed use/teach the word) that we elected to use the more common term “resonance structures”; other contexts (*e.g.*, in other countries) may use “mesoisomeric” more frequently, however. Similarly, “mesoisomeric hybrid” is also so uncommon that we use the term “resonance hybrid”. An expanded discussion of this idea and other instances of the use of “delocalization” instead of “resonance” are described in Appendix 3.1.

## 6.8 Theoretical Framework

#### *6.8.1.1 Learning outcomes*

Learning outcomes (LOs) are the knowledge, skills, and values that students *demonstrate* following a learning experience such as a module, section, course, program, or degree (Collis and Biggs, 1986; Brabrand and Dahl, 2009). Clearly identified LOs help focus the instructor’s and learners’ attention to specific outcomes (*i.e.*, what is demonstrably learned) instead of only the inputs (*i.e.*, what is taught). Clearly communicated LOs also allow students to track their own learning progress.

*Intended* Learning Outcomes (ILOs) are the knowledge, skills and values students are *expected* to be able demonstrate at the end of the course (*i.e.*, the intentions or inputs). Focused ILOs guide instruction to bring teaching strategies, learning activities (practice), and assessments into alignment (Figure 6.3). The learning environment and the learners’ characteristics are also essential to consider (see IPT section). In this study we use the terms LO to indicates the outcomes that students need in the future (*e.g.,* future courses) and ILOs to indicate efforts or objectives used to help students achieved those LOs.

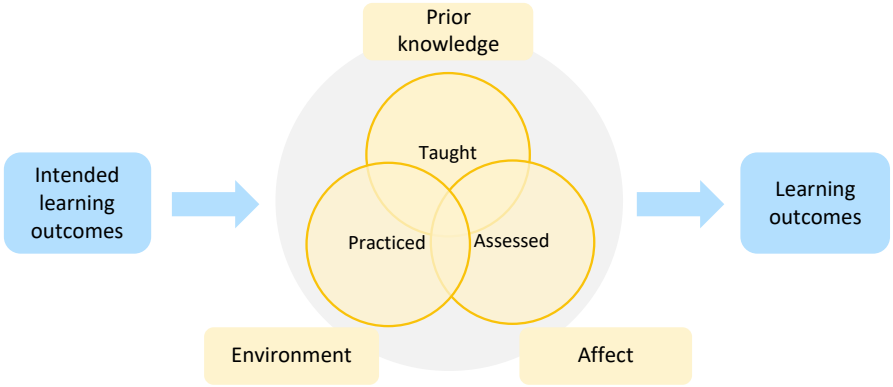



*Figure 6.3. ILOS can inform teaching strategies, learning activities, and assessment to achieve the desired LOs.*

The *taught* components convey the concepts in the ILOs to the learners (*e.g.*, videos, class notes, textbooks, and supplemental resources). Pedagogical choices are also a key component of the taught piece. The *practiced* part of the course involves the learning opportunities available to students, including class activities and problem sets. These activities can be used by educators, as formative assessments to enhance the learning activities and identify areas in which students need additional or different supports (Dixson *et al.*, 2016), or by the students to track their own knowledge and gain experience solving problems about the concepts. The *assessed* component looks at summative assessments (*e.g.*, midterm and final exams) that are used to make a decision such as giving a course grade, admission into a program or determine someone's abilities in a subject (Dixson *et al.*, 2016). An assessment can be both summative and formative; for example, midterm exams are used to give students grades and help the educator and students track the students' progress and guide further teaching and learning activities. For the purpose of this study, assessment refers to summative assessments, such as midterm and final exams.

To date, the only resonance-related LOs that have been proposed involve drawing the set of resonance structures and identifying the major contributors to the resonance hybrid (Betancourt-Pérez *et al.*, 2010); connections have not been made with other concepts such as molecular orbitals, relative stability/reactivity, or more advanced reactions.

#### *6.8.1.2 Information processing theory – Long-term memory storage*

This research is guided by the information processing theory (IPT) (Ausubel *et al.*, 1978; Sweller, 1999; Kalyuga *et al.*, 2003; Mayer and Moreno, 2003; Schunk, 2016). The theory states that an individual has two types of memory: working memory and long-term memory (LTM).

Long-term memory is where an individual stores their knowledge that can be retrieved at a later time. Information is stored as a network of concepts that are interconnected; recalling one concept can also activate the connected concepts. The information can be stored as declarative knowledge, which are facts and data obtained from events and experience, or procedural knowledge, which are the skills and procedures to perform a task (Cohen *et al.*, 1997; Gupta and Cohen, 2002). Both declarative and procedural knowledge are stored in LTM and are activated when prompted.

There are three factors that influence the storage of information in LTM: meaningfulness, elaboration, and organization. *Meaningfulness* means that the new knowledge must be presented in a meaningful way to the learner for the information to be integrated in LTM (Novak, 1993). *Elaboration* is the process adding the new pieces of information to knowledge already stored in LTM. Elaboration aids retrieval of information by providing alternate connections and adding more possible options for activation of pathways (Anderson, 1996), as well as adding more information to construct answers. *Organization* refers to how the information is integrated within the LTM network; a more organized LTM will facilitate the storage of information. Therefore, for a new piece of information to be stored in the LTM, it should be meaningful for the learner, elaborate on prior knowledge and be integrated in an organized network.

Storage problems will occur when a learner has no prior knowledge to elaborate the information from, rendering it meaningless. Conceptually meaningless information can be stored in the LTM; however, it may not be part of a network and may not be retrieved. Another problem with storage is that a learner may not elaborate properly and fail to create the connection to other concepts. This is why links between concepts should be explicitly mentioned when teaching, even when seeming obvious to an instructor.

Therefore, the optimal way to introduce information or concepts for learners to integrate it in LTM is to present it meaningfully and explicitly link the new information to prior knowledge. Instruction should follow a progression in which concepts learned build upon each other.

In this study, we examined delocalization concepts through a lens of IPT, specifically on how information is stored in LTM. We investigated whether the subject of delocalization is

presented in a way that is conducive for students to learn by investigating what and how the concepts are presented, what prior knowledge is cued in questions, how concepts are described, and what links are made between concepts.

## 6.9 Goals and research questions

The goals of this research were to (1) identify essential LOs related to delocalization that are needed to understand more advanced concepts in organic chemistry and related disciplines and (2) analyze how those intended LOs are taught, practiced, and assessed. Having defined LOs is intended to help educators design better teaching strategies, practices, and assessments. The research questions for this study are:

**RQ1:** What are the essential LOs related to electron delocalization, by the end of a two-semester sequence in organic chemistry?

**RQ2:** How are the proposed delocalization ILOs being taught, practiced, and assessed?

## 6.10 Methods

### 6.10.1 Data sources: textbooks and exams

One of the most commonly used teaching resources is textbooks; therefore, we used textbooks to gain an understanding of how delocalization concepts are being taught and practiced in organic chemistry courses. We analyzed seven introductory to Organic Chemistry, as they give a representation of what students are taught and intended to learn, providing insight into potential essential LOs (Solomons and Fryhle, 2000, 2011; Wade, 2010; Klein, 2012; Smith, 2014; Ogilvie *et al.*, 2018; Vollhardt and Schore, 2018). Six of the textbooks organize the information in the traditional functional group approach, while one textbook (Ogilvie *et al.*, 2018) used a mechanistic approach.

We also analyzed summative assessments from two main sources: (i) 51 different organic chemistry midterm and final exams at research-intensive universities in Canada, provided by the professors of those courses and (ii) information provided to us about the American Chemical Society's Organic Chemistry Practice exam, given by the ACS Exams Institute. The Practice Exam is designed to cover all topic areas equally; thus, the distribution of items on the Practice Exam does not necessarily reflect the distribution of items on a released exam.

### 6.10.2 Data Analysis

### 6.10.3 RQ1: What are the essential learning outcomes for delocalization concepts?

To identify the delocalization concepts that are taught, we performed several rounds of coding on seven introductory organic chemistry textbooks using a comparative analysis technique (Creswell, 2012). In the first round of coding, the first author identified chapters, or sections, in which delocalization was mentioned or required. We then used an inductive open-coding approach on those segments to produce codes that represented concepts being taught (Creswell, 2012). The codes were then aggregated and transformed into general categories taught in the textbooks. With these categories in hand, the authors went through several rounds of meetings to discuss and fine-tune the concepts until 33 fined grain ILOs were identified (Appendix 3.2). Specific coding was then employed; the textbooks were coded using the lens of the 33 ILOs and how the ILOs were taught, practiced, and assessed. The ILOs were then refined further, through discussion and expert validation to produce ten essential LOs.

#### *6.10.3.1 Validation of the LOs*

Seven experts were consulted, including two post-doctoral researchers and five professors with doctorates in organic chemistry. They were sent a survey in which they were asked to rank each of the 33 ILOs for Organic Chemistry II as essential, useful but not essential, not essential, or to be left for later courses. They were also given the opportunity to add any ILOs they felt were missing. For each ILO, a content validity ratio was calculated to determine the importance of that ILO (Zamanzadeh *et al.*, 2015). A content validity ratio over 0.8 indicates that the content of the item was considered essential by the majority of the experts. The survey also asked whether they used the terms resonance, delocalization, or both when teaching the concept to students and why. From these results and our coding of the textbooks, ten ILOs were determined as essential for the students to achieve.

### 6.10.4 RQ2: How is delocalization being taught, practiced, and assessed?

#### *6.10.4.1 Taught*

With all 33 ILOs in hand, we revisited the textbooks to determine how those ILOs were being taught (Solomons and Fryhle, 2000, 2011; Wade, 2010; Klein, 2012; Smith, 2014; Ogilvie *et al.*, 2018; Vollhardt and Schore, 2018). One, or multiple, of the 33 ILOs was assigned to each section that taught a delocalization concept (Figure 3). We also coded any inconsistencies or non-

specific language using a deductive coding approach (Creswell, 2012). When a paragraph of the textbook addressed delocalization, we coded how the subject was being taught and the language used (Figure 6.4). From this analysis we also identified common themes in how delocalization was being explained in the textbooks. An example of the coding can be found in Appendix 3.3.

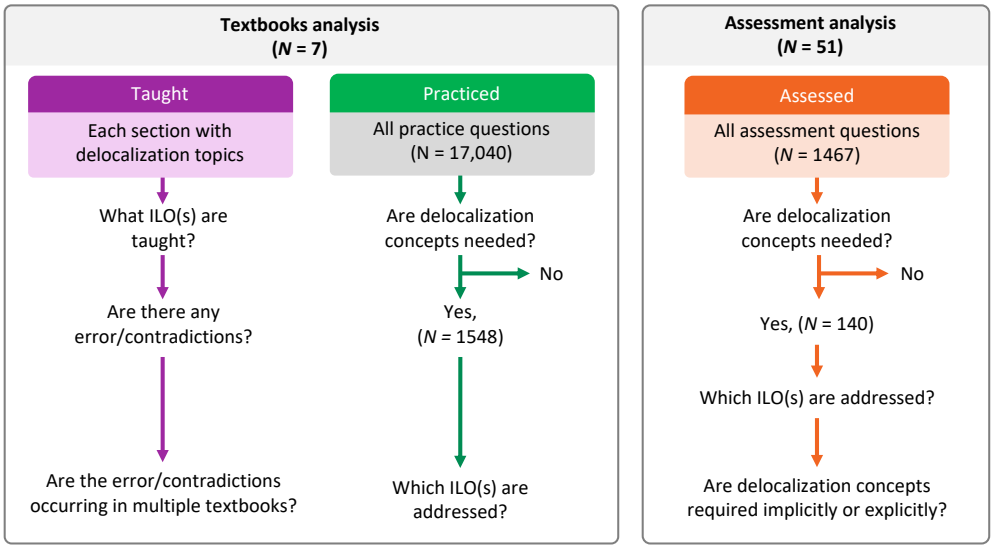



*Figure 6.4. Method to analyze the taught, practiced, and assessed aspects of delocalization concepts in organic chemistry.*

#### *6.10.4.2 Practiced*

We coded 17,040 end-of-chapter questions from the seven textbooks. Each question was coded as either requiring or not requiring delocalization concepts to answer. Questions that did not fully address any ILOs were coded accordingly. For the questions involving delocalization concepts, we identified the associated ILOs.

#### *6.10.4.3 Assessed*

Summative assessments (N = 51) were provided by seven professors, representing 25 different courses from 2011 to 2018, from research-intensive Canadian universities. All professors consented to have their assessments used in this research. The assessments included final examination and mid-term assessments and were analyzed for the presence of delocalization concepts. The majority (N = 47) only contained open ended questions, two had a mix of open-ended questions and multiple choices, and two assessments contained multiple-choice questions only.

Within the assessments, we coded each item (*i.e.*, question) as requiring delocalization concepts or not; if the item did require delocalization concepts, we identified which ILOs were being addressed. We also analyzed the items requiring delocalization concepts to determine whether the use of delocalization knowledge was implicitly or explicitly required. To be coded as explicitly required the use of delocalization knowledge, the question needed to state that resonance/delocalization or a related concept had to be part of the answer. For example, "Draw the mechanism of an electrophilic aromatic substitution, including all resonance structures" would be considered explicit, while "Which proton is more acidic? (In which one of the conjugate base can have delocalized electrons) would be considered implicit.

#### *6.10.4.4 Reliability of the coding*

To establish reliability of the coding, 15% of the textbook questions were randomly selected to be coded by a second rater. Four chapters from each textbook were also randomly selected for the second-rater to code, which amounted to 909 of the textbook questions. The raters obtained a 91% agreement and a Krippendorff □ of 0.7 before discussions, which is within the acceptable limit for reliability of coding (Krippendorff, 2004). The two coders also met and discussed the questions they had coded differently until agreement was reached.

## 6.11 Results and discussion

### 6.11.1 RQ1: Proposed Essential Learning Outcomes for delocalization

The ten LOs that arose from the analysis described in the methods section represent a combination of the authors' perspectives, analysis of more advanced contexts (*e.g.*, advanced organic chemistry, biochemistry), textbook analysis, and expert validation (Figure 6.5). The essential LOs are therefore meant to be representative of the skills and knowledge a student needs to have upon completing two semesters of organic chemistry. The LOs are outlined in such a way that they build upon each other.

\

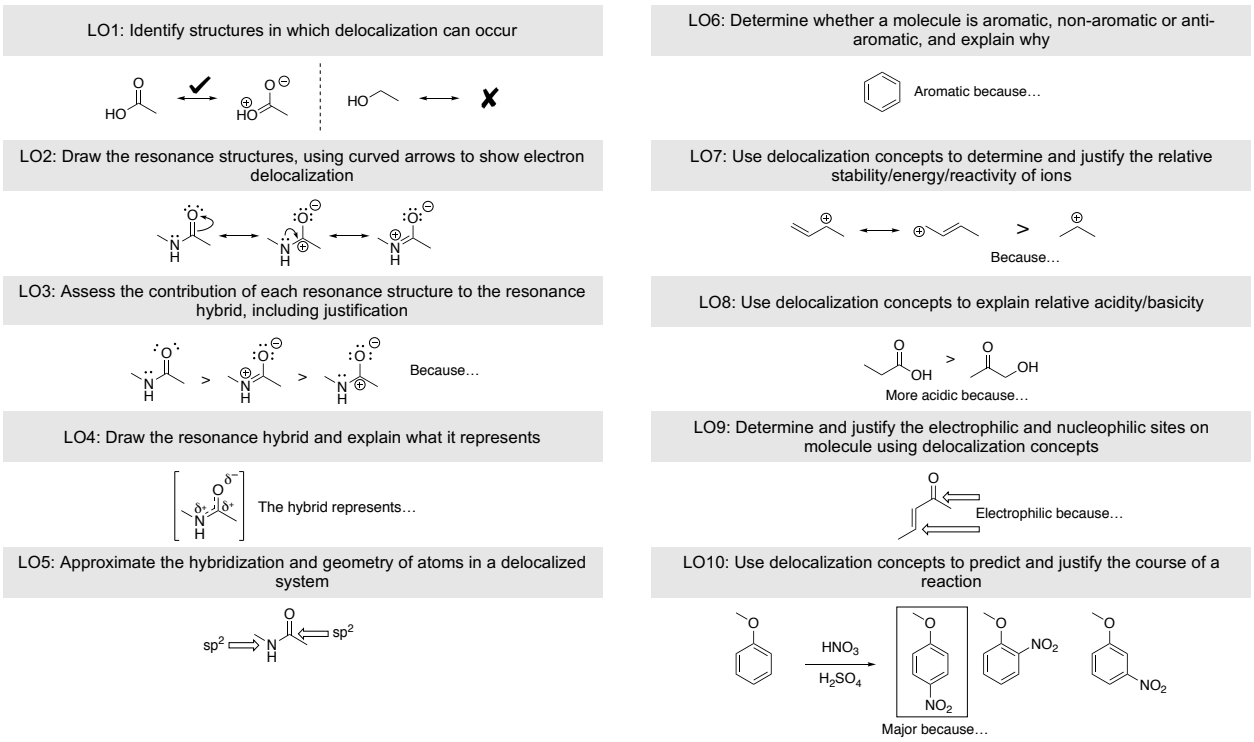



*Figure 6.5. The ten essentials LOs for the first two semesters of organic chemistry.*

The first LO identified is to *Identify structures in which delocalization can occur* ("*Identify*") which is a fundamental skill that is nested within many of the other LOs. The ability for students to quickly recognize whether delocalization is present within a molecule is necessary in many areas such as acid–base, stability, and properties of molecules. This ability also proves useful beyond organic chemistry, in disciplines such as biochemistry and physical chemistry.

The second LO is to *Draw the resonance structures, using curved arrows to show electron delocalization* ("*Draw*")*;* this LO requires skills with the electron-pushing formalism (EPF) and drawing Lewis structures. Previous research about resonance has shown a significant correlation between this LO and the students' overall success in organic chemistry courses (Betancourt-Pérez *et al.*, 2010).

The third LO is to *Assess the contribution of each resonance structure to the resonance hybrid, including justification* ("*Contribution*"). The goal is for students to rank the relative contribution of each resonance structure toward the resonance hybrid, leading eventually to an estimate of the reactivity of various sites.

LO four is to *Draw the resonance hybrid and explain what is represents* ("*Hybrid*"). We decided to include this LO although only five of the seven of the experts surveyed deemed it essential and it was only explained in four of the textbooks. We included this LO with the intent that competency with the hybrid representation could help dispel the notion that structures are alternating and give an alternate visualization of the concept of delocalization. This intent is supported by research findings that students better viewed resonance structures as an average representation of the structure after building their own representations and having a discussion about the pros and cons of representations (Kim *et al.*, 2019). Accordingly, teaching students how to draw the resonance hybrid and explaining what it represents could help students build a more scientifically accurate mental model of delocalization.

The fifth LO is to *Approximate the hybridization and geometry of atoms in a delocalized system* ("*Hybridization*"). This LO builds on the previous four outcomes since the learner must first identify that electrons are delocalized, and then approximate which structures are significant enough to affect the hybridization of an atom.

The sixth LO is to D*etermine whether a molecule is aromatic, non-aromatic, or anti-aromatic, and explain why* ("*Aromatic*"). Aromatic, non-aromatic, and anti-aromatic compounds have vastly different properties due to orbital, and electronic effects. Being able to conceptualize why a molecule is aromatic could allow students to better predict how the molecules will react.

The seventh LO, use delocalization concepts to determine and justify the relative stability/energy/reactivity of ions ("Stability/Reactivity"), and the eighth LO, Use delocalization concepts to determine the relative acidity/basicity of an atom ("Acid/Base") are similar. Delocalization is often the basis for a molecule's relative stability or acidity/basicity and using delocalization concepts has been previous identified as part of key LOs for acid–base chemistry (Stoyanovich *et al.*, 2015).

LO nine is to Determine the electrophilic and nucleophilic sites on a molecule using delocalization concepts as the justification ("Electrophilic/Nucleophilic"). Identifying nucleophilic and electrophilic sites is a key component to determining the reactivity of molecules; delocalization can affect the location of those sites.

The tenth LO is to *Use delocalization concepts to predict and justify the course of a reaction* ("*Reaction*"). This LO encompasses many of the previous LOs and applies the concepts in context. Delocalization can affect the regioselectivity of a reaction or whether a reaction occurs at all (*e.g.*, electrophilic aromatic substitution regioselectivity is dictated primarily by the extent of delocalization in the transition state leading to the arenium). Being able to identify how delocalization affects reactions stands to equip students with better skills to solve organic chemistry problems.

Of the ten LOs established, seven (LO3 – *Contribution*, LO4 – *Hybrid*, LO6 – *Aromatic*, LO7 – *Stability*, LO8 – *Acid/Base*, LO9 – *Electrophilic/Nucleophilic*, LO10 – *Reaction*) require a justification or explanation. Causal explanations address the reasons **why** a phenomenon occurs and allows learners, citizens, and scientists to go beyond a set of rules to understand, predict, and explain the cause-and-effect relationships that underpin phenomena (Talanquer and Pollard, 2010; Sevian and Talanquer, 2014; Cooper, 2015; Weinrich and Talanquer, 2016; Bodé *et al.*, 2019). Moreover, the US' National Research Council's *Framework* has emphasized the importance for students to develop causal reasoning skills when providing scientific explanations or arguments (National Research Council, 2012).

### 6.11.2 RQ2: How are the proposed delocalization intended LOs being taught, practiced, and assessed?

#### *6.11.2.1 Taught*

Of the ten proposed essential LOs, five were demonstrated in all of the textbooks analyzed (Figure 6.6) and four were underrepresented. ILO9 (*Electrophilic/Nucleophilic*) was taught in only one of the textbooks (Ogilvie *et al.*, 2018), leaving it to students to make the connection between delocalization concepts and electrophilic/nucleophilic sites on a molecule. Similarly, ILO1 (*Identify*) was only explicitly demonstrated in three of the textbooks (Klein, 2012; Smith, 2014; Ogilvie *et al.*, 2018). Similarly, only a few textbooks explained explicitly how to A*pproximate the hybridization and geometry of the atoms involved in delocalization* (ILO5) (Wade, 2010; Klein, 2012; Smith, 2014). *Draw the resonance hybrid and explain what it represents* (ILO4) was also under-represented. The resonance hybrid was shown in all of the textbooks, with four of them explaining how to draw the hybrid and what the dashed bonds represent (Solomons and Fryhle, 2000, 2011; Smith, 2014; Ogilvie *et al.*, 2018). Two of the textbooks explained that

the hybrid is a representation of all resonance structures but did not explain how to draw the hybrid or what the dashed bonds represent (Wade, 2010; Vollhardt and Schore, 2018). The remaining textbook mentioned the hybrid is the "true" representation but did not show an image of the hybrid (Klein, 2012). Explaining the hybrid structure and its components could help enhance students' representational competency and mental models of delocalization concepts.

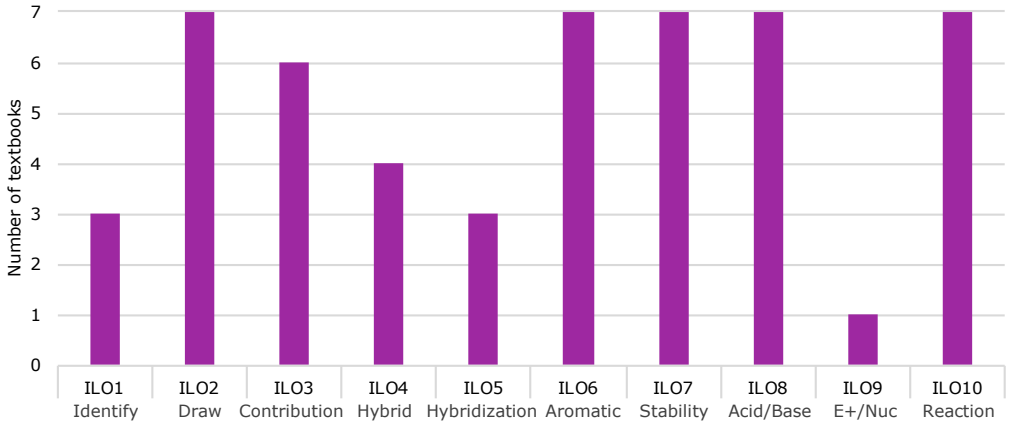



*Figure 6.6. Occurrence of the ten essential ILOs in the textbooks (N = 7).*

Within the textbooks analyzed, the ILOs were addressed in two main areas, the beginning and the middle of the textbooks (Figure 6). ILOs 2 through 4 (*Draw*, *Contribution*, and *Hybrid*) were shown within the first two chapters of each book. ILO7 (*Stability*) and ILO8 (*Acid/Base*) were shown in the following few chapters, when introducing the acid/base and substitution/elimination reactions, respectively. ILO6 (*Aromatic*) and ILO10 (*Reaction*) were taught in the middle of the textbooks when conjugation and aromaticity were introduced. ILO5 (*Hybridization*) appeared in different places, in that two textbooks taught the concept in the first two chapters, while the remainder addressed it when introducing aromaticity. This placement could be occurring because the orbitals, hybridization and geometry is taught after delocalization in four of the textbooks (Solomons and Fryhle, 2000, 2011; Wade, 2010; Vollhardt and Schore, 2018) while the other three teach delocalization after having gone over hybridization. The order in which those concepts are taught makes a difference of over ten chapters between the textbooks analyzed in when the concept of hybridization in terms of delocalization is taught.

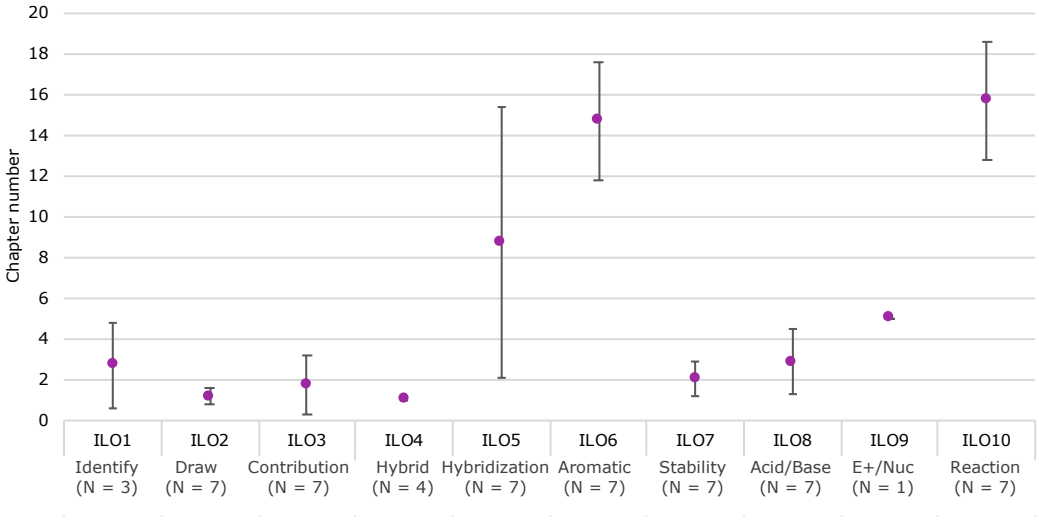



*Figure 6.7. The average location in the textbooks where the ILOs are taught. Error bars denote the standard deviation.*

#### *6.11.2.2 Gaps are present in textbooks related to delocalization concepts*

The textbooks analyzed all mention the fact that resonance structures are not real and not alternating, an important idea to emphasize that previous research indicating that some students interpreted resonance structures as oscillating forms (Taber, 2002). However, beyond a first mention, that message is not repeated throughout the textbooks (in explanations or questions), which poses a concern for student knowledge development of a key concept.

Another potential issue involves the slightly differing meaning of the electron pushing formalism symbolism for reactions and resonance contribution (Figure 7). For example, in the reaction shown in Figure 6.8, there are curved arrows that show electron movement for a reaction and curved arrows that show electron delocalization. This symbolism could add to the idea that student have that resonance structures are alternating, and the electrons are moving. This potential confusing could be mitigated by clearly explaining the different ways the curved arrows are used and by using and explaining the hybrid (ILO4).

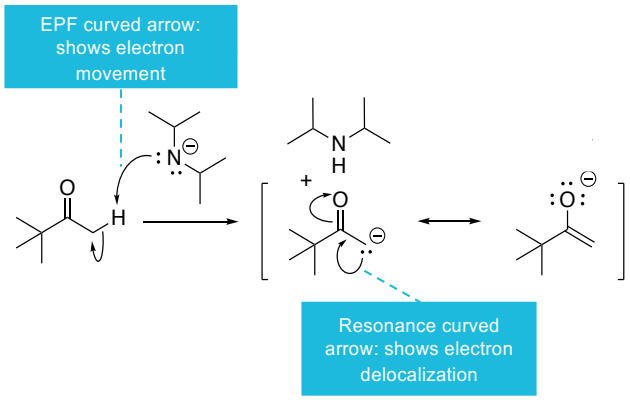



*Figure 6.8. Two different meanings and uses of curved arrows.*

Another misalignment with the concept of delocalization is that resonance structures are often labelled as "stable" and describing relative stability of resonance structures could be misleading because resonance structures do not exist as discreet entities. Instead, referring to the individual structures as contributors (*i.e.*, major contributor, minor contributor) avoids giving a real property (stability) to a non-real, as do some textbooks (Klein, 2012; Smith, 2014; Ogilvie *et al.*, 2018).

All the textbooks analyzed used the term resonance more often than the word delocalization. When asked, experts reported they use the terms resonance and delocalization interchangeably. One expert stated: "*Resonance is a term that is too prevalent, so we need to teach it. Delocalization is also useful and arguably a better term pedagogically."* This quote shows that resonance is an integrated part of the organic chemistry culture, despite being the less precise term for the phenomenon. We analyzed two more advanced textbooks (Fleming, 2009; Clayden *et al.*, 2012) and found that they refer to the concept as delocalization; they also mentioned that chemists will often use the word resonance. Clayden et al. state in the textbook that they avoid the use of the term "resonance" and only uses "delocalization" to avoid confusion:

"How to describe delocalization? What word should be used to describe delocalization is a vexed question. Terms such as resonance, mesomerism, conjugation, and delocalization are only a few of the ones you will find in books. You will already have noticed that we're avoiding resonance because it carries the suggestion that the structure is somehow oscillating between localized structures." (Clayden, Greeves, and Warren, 2012, p. 145). Explicitly teaching these

limitations and using precise terminology could help avoid confusion and help learners integrate the concept.

#### *6.11.2.3 Causal reasoning is infrequently connected with delocalization concepts*

When designing the essential LOs, we ensured that seven of the ten involved causal reasoning. Research has shown that building causal connection through explanations or argumentation facilitate better conceptualization of a scientific concept (Arterhan and Schwarz, 2007). Building an explanation or argument requires the students to use evidence of the scientific concept, articulate their knowledge and persuade others (Berland and Reiser, 2009; Cooper, 2015). The students need to build their knowledge structures to include the reason a phenomenon occurs, understand the underlying concept, and transfer their knowledge to new scenarios. Most LOs herein include causal reasoning components so that learners can explain the reasons delocalization occurs and how delocalization affects molecules' stability, reactivity, and properties.

In the textbooks analyzed, causal reasoning was not usually included when teaching delocalization; rule-based reasoning was used instead (Kraft *et al.*, 2010). As one example, the relative contributions to a resonance hybrid are taught as a set of rules: the highest maximum number of bonds to an atom or maximum number of atoms with a full octet, minimum number of charges, charge separation, *etc*. Four textbooks used chemical stability as the reasoning being the contribution of the hybrid (Solomons and Fryhle, 2000, 2011; Wade, 2010; Vollhardt and Schore, 2018). Only one of the textbooks (Klein, 2012) gave reasons as to why a contributor is major (*i.e.*, the causal links).

Another example of a common organic chemistry rule related to delocalization is Hückel's rule. Hückel's rule states: "Monocyclic planar (or almost planar) systems of trigonally (or sometimes diagonally) hybridized atoms that contain $(4n + 2)$ π-electrons (where $n$ is a non-negative integer) will exhibit aromatic character." (McNaught and Wilkinson, 1997) This rule does not explain why these factors are required for aromaticity, such as planarity and π-orbital alignment. By incorporating causal reasoning into the LOs and educational design, students stand to be better positioned to move from rote memorization toward more meaningful learning.

#### *6.11.2.4 Delocalization was not explicitly linked to other chemical concepts*

Delocalization affects many aspects of a molecule such as stability, chemical properties, and reactivity. However, our textbook analysis found that these concepts were not always related explicitly to the concepts of delocalization. The connections between the concepts and delocalization were not explicit and therefore had to be made by the learner themselves.

For example, ILO9 (*Electrophilic/Nucleophilic*) was only explicitly taught in one of the textbooks (Ogilvie *et al.*, 2018). Analysing structures for electron delocalization can reveal areas of high and low electron density, which can be used to identify nucleophilic and electrophilic sites. Delocalization is taught before the concepts of nucleophiles and electrophiles, and when nucleophiles and electrophiles are taught, the effect of delocalization is not explained. The link between delocalization and electron density may be obvious to experts but this connection is often only made implicitly in textbooks and so may never become apparent to learners.

Similarly, ILO5 (*Hybridization*) is not explicitly explained in four of the textbooks. One of the three that did explain this LO mentioned it in a solved practice question, not as part of the main text (Wade, 2010). Ogilvie et al. (2018) gave a brief explanation in Chapter 5 (Organic Reaction Mechanisms: Using Curved Arrows to Analyze Reaction Mechanisms) (p. 220) although delocalization was first introduced in Chapter 1: (Carbon and Its Compounds). Klein (2012) gave a detailed explanation of how to identify the hybridization and geometry of the atoms in an amide functional group. Greater explanations around the concepts of hybridization are likely needed.

ILO8 (*Acid/Base*) and ILO7 (*Stability*) were explained only very briefly in most of the textbooks. The effect of delocalization on a molecule's stability was briefly explained in the first chapter (second chapter in Klein (2010)); however, the concept was not revisited when describing ions. Two of the textbooks explained the reasons why delocalization affects acidity/basicity (Smith, 2014; Ogilvie *et al.*, 2018). The others make broad statement such as: "*In this case, the charge is delocalized over both oxygen atoms. Such negative charge will be more stable than a negative charge localized on one oxygen"* (Klein, 2012). This explanation will promote rule-based reasoning and does not explain **why** delocalization stabilizes ions.

Our analysis shows that the concept of delocalization is sometimes overlooked in key areas, such as electron density, hybridization, acid–base chemistry, and stability. The concepts are taught in silos without an explicit connection being made, which could lead to difficulty storing the information in LTM due to the lack of the explicit links.

### 6.11.3 Practiced

Practice problems and feedback help students build organic chemistry skills can help build metacognitive skills, both ideally promoting meaningful learning (Grove and Lowery Bretz, 2012). Practice questions help students store information in LTM by repeatably elaborating pieces of information and organizing the information into a network (Gupta and Cohen, 2002).

We analyzed the 17,040 end-of-chapter questions from the seven textbooks; of those, 9% (N = 1548) related to delocalization (Figure 6.9). Many questions included sub questions (*i.e.*, a, b, c, *etc.*). The questions varied from multiple choice (*e.g.*, select the most stable molecule) to constructed response questions (*e.g.*, draw the mechanism of this reaction).

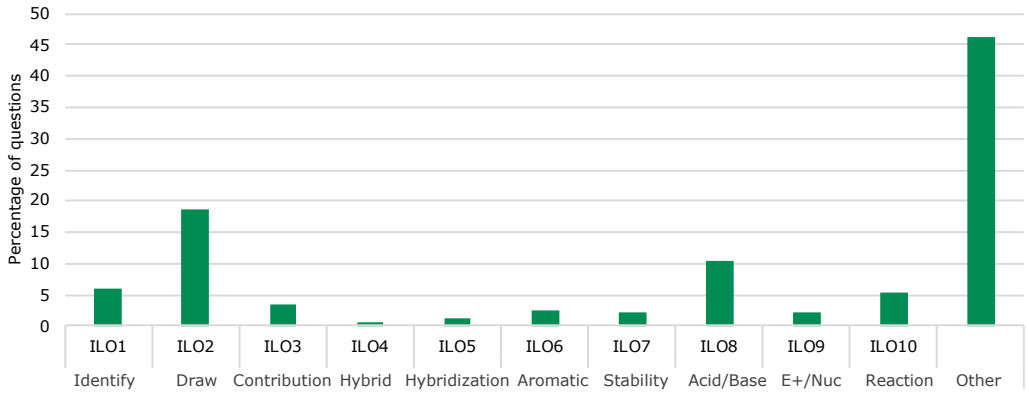



*Figure 6.9. Percentage of questions related to the LOs from all delocalization questions (N = 1548). "Other" represents any delocalization questions that do not fully address one of the 10 essential LOs.*

Of the ten essentials intended LOs, ILO2 (*Draw*) was the most commonly practiced with 15% of all delocalization questions. The remaining intended ILOs were not present as often with ILO5 (*Hybridization*), ILO7 (*Stability*) and ILO9 (*Electrophilic/Nucleophilic*) each representing less than 2% of the questions about delocalization.

The least practiced essential ILO was ILO4 (*Hybrid*), with only 0.7% of all the delocalization questions being related to it. Only three of the seven textbooks had practice questions related to the resonance hybrid (Solomons and Fryhle, 2011; Smith, 2014; Vollhardt and Schore, 2018).

The "Other" category occurred more often than all the other intended ILOs since most questions did not fully address the essential ILOs and instead related to the ones considered non-essential (Figure 8). Many of the questions did address parts of the essential ILOs but did not require justifications or explanations. These were coded as partially assessing the ILOs and were added to the "Other" category along with questions related to delocalization by not within the ten essential LOs. For example, the "Other" category includes *Predict the product of an aromatic electrophilic substitution reaction*. Solving electrophilic aromatic substitution reactions involve the knowledge of delocalization but can be readily answered by memorizing the directing groups' effects without knowing the reasons for those effects. Therefore, simply predicting the product does not fully address LO10 (reaction) since it lacks the justification component. *Predict the product of an electrophilic substitution reaction* represented 20% of all delocalization questions while *Draw the electrophilic aromatic substitution mechanism with all structures representing delocalization* occurred only in 2.7% of the questions. Similarly, the LO *Use Hückel's rule to determine if a molecule is aromatic, anti-aromatic or non-aromatic* was found in 2.5% of all delocalization questions, while justifying aromaticity using other parameters was found in 1.1% of questions. Appendix 3.2 shows the distribution of questions to all of the 33 ILOs.

While the "non-essential" LOs could be important, having over 50% of the practice questions addressing them identifies a concern about how questions are designed and chosen. Shifting the questions proportion to align with the more the essential LOs could promote better understanding of delocalization.

#### *6.11.3.1 Many chapters separate teaching the concept and using it in context*

Delocalization is introduced in the first chapter in six of the seven textbooks alongside how to draw a Lewis structure and how to draw the curved arrows to demonstrate electron movement (Klein (2012) introduces delocalization in the second chapter). Since students lack the necessary prior knowledge early in the course—*i.e.,* how to draw Lewis structures and curved arrows—they may struggle to incorporate delocalization concepts in their long-term memory. Also, students are not required to immediately use delocalization concepts after they are taught, meaning that several chapters go by before the subject of delocalization is broached again. On average, 13 chapters separate the introduction and use of delocalization concepts. Specifically,

Chapters 3 to 13 had very few delocalization related questions (Figure 6.10). Since delocalization concepts are not used immediately, students may not incorporate the knowledge into their LTM because they may not identify these concepts as being important or link them with relevant knowledge (*e.g.*, nucleophilicity and electrophilicity).

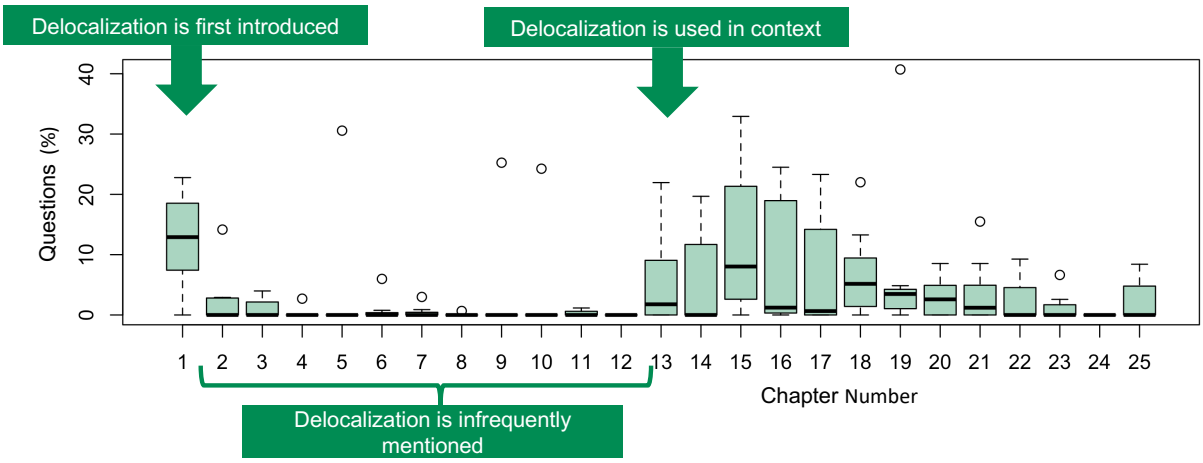



*Figure 6.10. Distribution of delocalization question (%) in each chapter from the seven textbooks. N = 1548*

The first-time delocalization is used in context is during the acid–base chapter to talk about acidity/basicity, which appears six chapters after delocalization has been taught, on average. We propose introducing delocalization when it is needed, specifically at this point and as has been done in previous work (Stoyanovich *et al.*, 2015). This placement would mean introducing delocalization when introducing acid–base chemistry instead of when introducing Lewis structures. This placement would ideally help the students understand why delocalization concepts are important. Teaching delocalization later would also allow the students to have gained the prior knowledge required (drawing and interpreting Lewis structures and electron-pushing arrows) to integrate the new knowledge of delocalization.

Delocalization affects reactivity, stability, and properties of a molecule; therefore, we anticipated seeing questions related to delocalization in chapters about structure (including ions and radicals), conjugation, aromaticity, and reactions (*e.g.*, acid–base, elimination, substitution, and electrophilic aromatic substitution). Our analysis found that sections on structure, conjugated systems, aromatic chemistry, and electrophilic aromatic substitution reactions did have a higher percentage of delocalization questions, while acid–base chemistry, elimination and

substitution reactions, and radicals had a low number of delocalization questions (Figure 6.11). Only 1.4% of all delocalization questions were related to cation stability with delocalization, a subject that is highly important for substitution and elimination reactions.

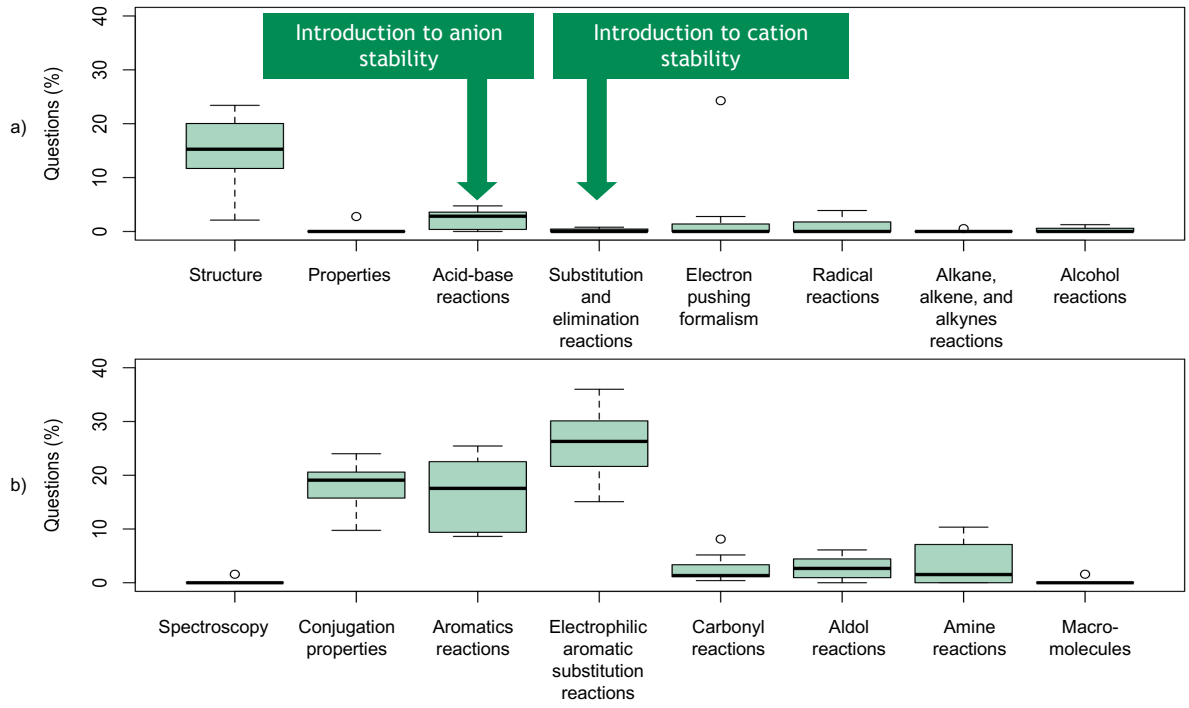



*Figure 6.11. The percentage of delocalization questions (N=1548) by subject. The top graph (a) is the traditional curriculum for Organic Chemistry I and the bottom graph (b) is the traditional curriculum for Organic Chemistry II.*

### 6.11.4 Assessed

Assessments are used to assess the progress toward the intended LOs and should be aligned with how a concept is being taught and practiced, as well as the intended LOs themselves.

#### *6.11.4.1 Delocalization questions on summative assessments*

On average, the concept of delocalization was required for 10% of summative assessments on both mid-term and final exams of Organic Chemistry I and II courses. The value of delocalization concepts on assessment seems well aligned with the 9% of practice questions, in textbooks, related to the concepts.

In terms of the breakdown by intended LO, LO4 (*Hybrid*) was addressed only in one of the questions analyzed (Figure 6.12). This could be due to the convention that practicing organic chemists use the resonance structures representations as opposed to the resonance hybrid.

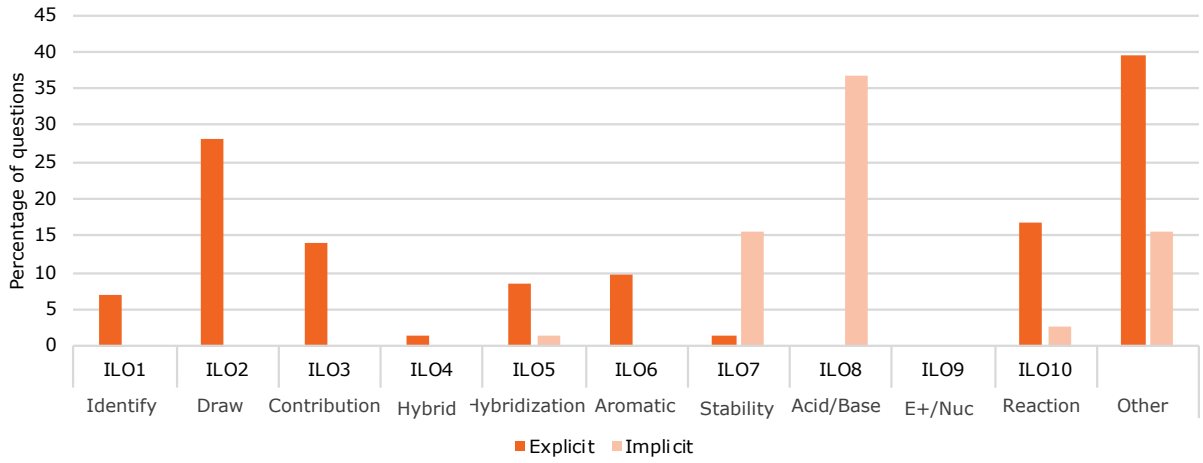



*Figure 6.12. Percent of delocalization questions on assessments (N = 143) that are explicit (dark orange) and implicit (pale orange). The "Other" Category includes all questions that partially assess the LOs and any questions assessing "non-essential" ILOs.*

LO9 (*Electrophilic/Nucleophilic*) was completely absent from the assessments analyzed, but 6% of the questions partially addressed this LO. For example, a question asking students to classify substituents as *ortho, meta* or *para* directing does not fully address the LO, because it does not require the delocalization justification and could be simply memorized. Similarly, questions related to the ILO: *Predict the electrophilic aromatic substitution product* requires an understanding of the regioselectivity but may be answered without having the skill and knowledge of LO9 (*Electrophilic/Nucleophilic)*.

LO 10: *Use delocalization concepts to predict and explain the course of a reaction* represented 10% of the assessment questions; however, "Predict the product" questions, which do not require justification or drawing of the resonance structures, that involved delocalization represented 16% of the questions. While students may have the skills to use delocalization to predict a reaction, without the justification or drawing of the structures, LO 10 cannot be fully assessed.

There was a similar number of questions in both Organic Chemistry I and Organic Chemistry II exams (67 and 76 respectively) but there was a difference in the nature of LOs that

were assessed (Figure 6.13). Assessments in Organic Chemistry I focussed more on drawing resonance structures and assessing the structures (LO2 – *Draw*, LO3 – *Contribution*, LO5 – *Hybridization*, LO7 – *Stability* and LO8 – *Acid/Base*) while Organic Chemistry II assessments focused more on reactions using delocalization (LO6 – *Aromatic*, LO7 – *Stability*, LO8 – *Acid/Base* and LO10 – *Reaction*).

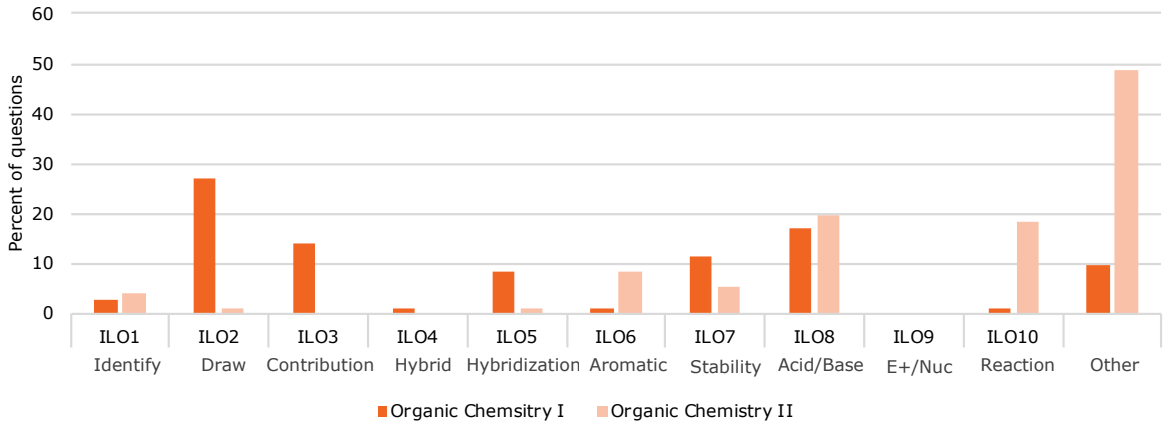



*Figure 6.13.Percent of delocalization question (N = 143) related to each ILO on professors' assessments (N = 51) in Organic Chemistry I (dark orange) and Organic Chemistry II (pale orange)*

We also reached out to ACS Exams to determine the weight of delocalization on the ACS Organic Chemistry Practice Exam. One item (out of 50) on the Organic Chemistry Practice Exam explicitly assessed delocalization. Other questions may indirectly assess delocalization; however, no in-depth analysis was done.

### 6.11.5 Implicit vs explicit delocalization questions

Questions that implicitly involved delocalization concepts had significantly greater value ($t(49) = 6.91$, $p < 0.001$) than explicit delocalization questions on the summative assessments. Questions that implicitly involve delocalization concepts do not state that the use of delocalization is required, but answering the question requires the student to use knowledge of delocalization. Explicit delocalization questions clearly state the requirement for using delocalization/resonance concepts. Both implicit and explicit questions are useful and assess different aspect of the concept.

Some LOs were almost always explicit (LO1 – *Identify*, LO2 – *Draw*, LO3 – *Contribution,* LO5 – *Hybridization* and LO6 – *Aromatic*). These LOs deal with the structures and property of

delocalization, while the remaining LOs are typically used in a context. For example, all of the delocalization question about LO8 (*Acid/Base*) were implicit. The most common acid–base question required that students compare two protons in different molecular environments with one of them having delocalized electrons in their conjugate base (Figure 12). Similarly, LO7 (*Stability*) was often implicit. LO10 (*Reaction)* was both implicit and explicit; the majority of the questions on LO10 explicitly stated that resonance structures should be drawn for full marks. However, there were also several questions in which a mechanism was all that was required, and that delocalization was implicit.

## 6.12 Conclusions

Through this study, we proposed ten essential learning outcomes (LOs) for the subject of delocalization (Figure 4), designed based on key concepts related to delocalization and considering ways in which delocalization is used in more advanced contexts (*e.g.*, identifying sites of high/low electron density). Identifying explicit essential LOs can help align curricular and instructional practices, including assessment.

We conducted a textbook analysis using those ten intended LOs and found that they were not always the focus of how delocalization was taught, practiced, and assessed. Some of the intended LOs were underrepresented in the textbook explanations or the practice questions (*e.g.*, end-of-chapter questions). In particular, underrepresented LOs in explanations were LO1 (*Identify*), LO4 (*Hybrid*), LO5 (*Hybridization*), and LO9 (*Electrophilic/Nucleophilic*). In textbook practice questions and summative assessments, only LO2 (*Draw*) and LO8 (*Acid/base*) were well-represented; questions related to the other essential ILOs were infrequently asked. There were many delocalization questions related to non-essential ILOs (Figure 8). Moreover, there was a large gap in most textbooks between the chapter when delocalization concepts were first explained and the chapter in which the concepts were used in context (10 chapters, on average), posing likely learning issues for students in seeing the relevance of delocalization initially and connecting concepts later (Figure 9).

We found potential areas of confusion in the way delocalization concepts were explained, which could be mitigated by more detailed and explicit explanations, including: (i) curved arrows can be used to describe electron movement in reactions (EPF) and to explain the differences

between resonance structures (not electron movement *per se* since the electrons are delocalized and structures are not alternating), (ii) resonance structures were frequently described in terms of relative stability; however, since these structures do not exist discretely, it would be more appropriate to describe them in terms of their relative contribution to the resonance hybrid, and (iii) the term "resonance" has been interpreted by students as structures that alternate; instead, the term "delocalization" could be more precise and help students develop more accurate mental model of delocalization.

The focus in most questions (in the textbooks and on the exams) and explanations (in the textbook) was on rule-based explanations as opposed to causal explanations. Summative assessments questions (*e.g.*, final exams) followed a similar pattern as the practice questions. Teaching and assessing by including causal reasoning can help students consolidate ideas and provides opportunities to practice building explanations and arguments from evidence.

## 6.13 Limitations

Textbooks are a common component of chemistry courses and were used in this study to capture how delocalization concepts are taught; however, they do not represent the entire educational picture. Instructors may not follow exactly the layout of the textbook, use supplementary material, or may not use a textbook at all. Nevertheless, textbooks reflect how authors (and textbook adopters) believe chemistry should be taught and therefore provide one lens into how a concept is being taught. Another limitation of this study is that no student data was analyzed; therefore, we cannot make claims about students' achievement of the essential.

## 6.14 Implications for teaching and learning

The ten evidence based essential LOs can be used by educators in their courses to guide instruction and assessments. The essential LOs will ideally also be shared with students to communicate the expectations and skills they should master and help them focus their studying. The LOs could also be used in the design of new curriculum in a backward design method (Richards, 2013). The method first identifies the goals or skills that need to be attained by the end of the curriculum and then selecting the appropriate methods for instructions and assessment. These LOs can act as the goals of the instruction and help guide curriculum design and learning opportunities for the university organic chemistry sequence.

Educators could review their own formative and summative assessment questions with the essential LOs in mind to identify areas where they may wish to adjust their own instruction and assessments.

## 6.15 Implications for research

New research could be guided by the ten essential LOs established, and how delocalization is taught, practiced, and assessed.

The LOs proposed also build on each other; therefore, investigating how students' progress through the LOs could be investigated. The LOs could be used as the basis for a learning progression about the subject of delocalization.

We have introduced an approach to look at how a concept is being taught and found evidence-based ILOs that can guide instruction. This approach could be used on other chemistry concept to identify essential LOs.

## 6.16 Acknowledgements

We thank the following people who provided us with past assessments and/or expert advice: Dr. André Beauchemin, Dr. Amanda Bongers, Dr. Kathy-Sarah Focsaneanu, Dr. Stephen MacNeil, Dr. Katherine McGilvray, Dr. Barb Morra, Dr. Leanne Racicot, Dr. Effiette Sauer, Dr. Jaclyn Stewart, and Dr. Alison Thompson. We also thank Keith Lapierre for his work on the inter-rater reliability.

## 6.17 Submitted contribution: How are the learning outcomes being achieved

*Ten essential delocalization learning outcomes: How well are they achieved?*

Myriam S. Carle, Roméo El Issa, Nicolas Pilote, and Alison B. Flynn

**Abstract:** OBJECTIVE: Delocalization (resonance) is a concept in organic chemistry that influences the chemical reactivity, activity, structure, and physical properties of molecules. However, the concept has proven challenging for students. The goal of the present study was to investigate to what extent ten essential delocalization learning outcomes (LOs) were achieved by students, how students use and reason about delocalization as well as the connections between the LOs. The goal is to discover where and how students may be struggling when answering delocalization-related exam questions and uncover potential barriers to learning delocalization.

METHODS: We analyzed students' responses (N = 3787) on twelve exam questions related to seven of the ten LOs for the degree of achievement, common errors, and scientific reasoning.

RESULTS: The achievement on the LOs was variable. We report types of errors and strategies used, the errors are primarily related to drawing resonance structures or the resonance. Six key findings emerged from the analysis: (1) the majority of answers had few (<10%) representational errors (2) in an implicit question where delocalization or inductive effect concepts could be used to justify a response, half the students used delocalization concepts, (3) delocalization was used in 10–20% of answers when relevant but not prompted or required, (4) strategies that helped students reason with the representations (*i.e.*, drawing out electrons or expanding a structure) were correlated with higher achievement of the LOs, (5) students' reasoning aligned with course expectations, and (6) students who achieved later LOs typically (60–95%) also achieved LO1 and LO2 (*Identify that electron delocalization is relevant, Draw resonance structures*).

CONCLUSIONS: The findings have implications on how students achieve the LOs and suggest ways educators can better support learners with the tools to achieve the LOs.

IMPLICATIONS: The findings from this work could be used to design and evaluate new teaching techniques or materials, including scaffolding concepts. Further investigations could

lead to a deeper understanding of students' mental models and thought processes related to delocalization concepts.

**Reproduced with permission from**: Carle, M. S., El Issa, R. J., Pilote, N., & Flynn, A. B. (2021). Ten Essential Delocalization Learning Outcomes: How well are they achieved? *Journal of Chemical Education.* Revisions requested. Available at https://doi.org/10.33774/chemrxiv-2021-c12b9-v2 D O I: 10.33774/chemrxiv-2021-c12b9-v2

**Author contribution:** REI performed the first round of coding on Question 3 and provided a preliminary code book for LO10 (Reaction) and LO2 (Draw – within a mechanism). NP performed the first round of coding on Question 2 and provided a preliminary code book for LO2 (Draw), LO7 – Stability, and LO8 (Acid—base). MSC performed the first round of coding for all other questions. MSC used all preliminary data to create a finalized code book and re-coded all exams. MSC performed the data analysis and drafted the manuscript, making edits based on feedback. The authors discussed all phases of the study. ABF provided guidance and feedback on all stages of the project.

## 6.18 Introduction

Delocalization, or resonance, is an integral part of organic chemistry, affecting the structure, properties, activity, and reactivity of molecules and is present in most medicines, biological structures, materials, and other compounds. Educators have described that the subject of delocalization is difficult to learn and that students struggle with the concept (Duis, 2011). Empirical studies have revealed several incorrect ideas about resonance, such as the view that resonance structures are alternating or that resonance structures exist as an equilibrium (Taber, 2002; Xue and Stains, 2020), or may not even think to use delocalization and would have to be prompted to do so (George M. Bodner and Domin, 2000; Finkenstaedt-Quinn *et al.*, 2020; Petterson *et al.*, 2020; Watts *et al.*, 2020). One intervention focused on building and exploring the representations in delocalization (*i.e.*, hybrid and resonance structures) showed improvement in students' conceptions of the resonance structures (Kim *et al.*, 2019). Similarly, students taught by an instructor who focused on the meaning and limitations of resonance structures had a higher conceptual understanding of resonance compared to students in a course setting that emphasized identifying/drawing only (Taber, 2002; Xue and Stains, 2020).

We recently proposed ten essential learning outcomes (LOs) about delocalization that a student should be able to demonstrate by the end of an organic chemistry course sequence (Figure 6.14), which address concepts of structure, properties, and reactivity (Carle and Flynn, 2020). The LOs are based on interviews with faculty, textbook analyses, analyses of the knowledge and skills required in future courses, and past literature (Betancourt-Pérez *et al.*, 2010).

The goal of this research was to investigate how the ten essential delocalization LOs have been achieved on summative examinations by students in Organic Chemistry I and II courses at one institution. In doing so, we sought to connect existing literature of students' difficulties learning resonance concepts described above with the clearly defined expectations (intended LOs), working with a larger sample size. We also investigated their strategies and explored the ways in which students may be struggling in their learning, to uncover potential barriers to learning more advanced delocalization or chemistry concepts.

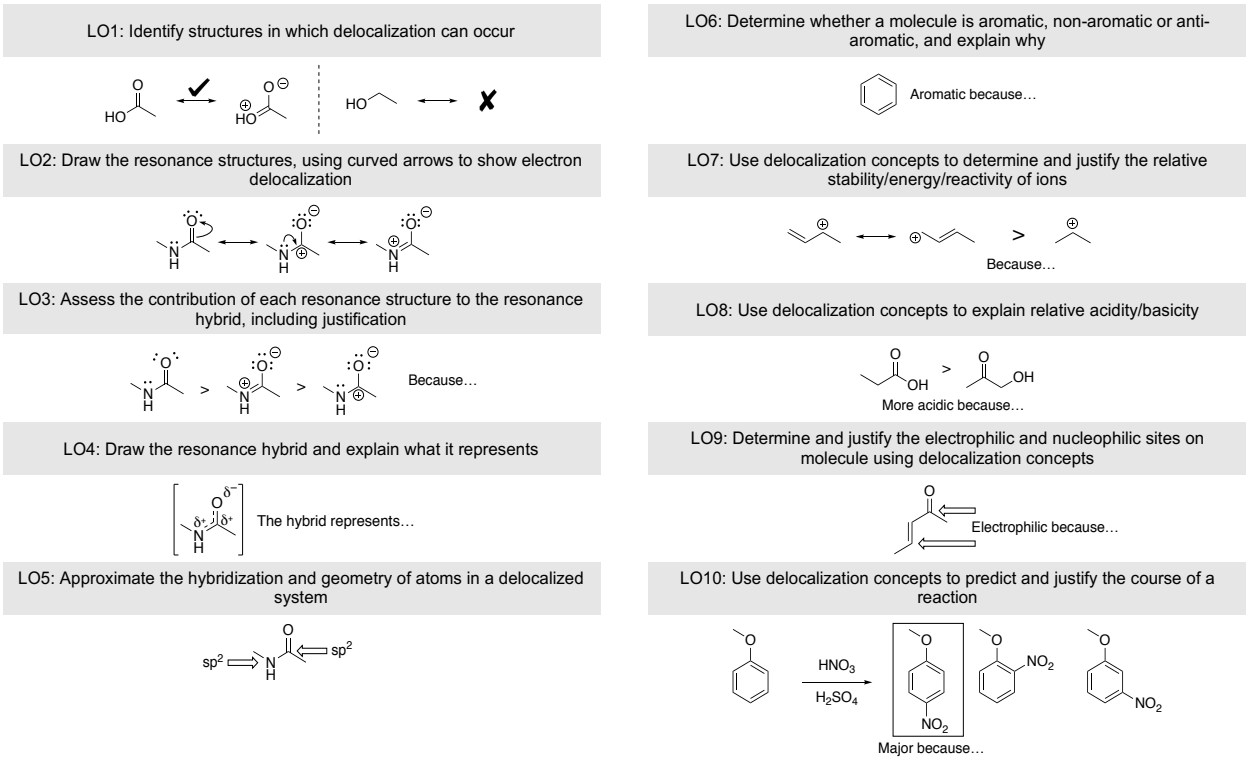



*Figure 6.14. The ten essential LOs for delocalization. Reproduced with permission from Carle and Flynn, 2020*

Herein, we use the term delocalization to represent the concept, except when talking about resonance structures or the resonance hybrid, or when quoting question statements or students.

## 6.19 Research questions

We explored the following three research questions to achieve the project's goal:

RQ1: What is the relative achievement rate on the delocalization related LOs, as demonstrated by summative assessments?

RQ2: What are the common strategies and errors on delocalization-related exam responses?

RQ3: What is the sophistication of arguments for delocalization-related exam questions that require justification?

## 6.20 Theoretical frameworks

We used three frameworks: LOs (described above), modern information processing theory, and a modes of reasoning framework (Sevian and Talanquer, 2014).

### 6.20.1 Modern information processing theory

This research is guided by modern information processing theory (Ausubel *et al.*, 1978; Sweller, 1999; Kalyuga *et al.*, 2003; Mayer and Moreno, 2003; Schunk, 2016). When faced with information, we process information in working memory, which is short-term and can only hold a limited amount of information. Memories are stored in long-term memory within an interconnected network, called schema, with links between the concepts.

Once in the long-term memory, information can be recalled if cued. A specific concept can activate a schema where all connected information can be accessed in the working memory. However, some information may not be recalled because of inference (*i.e.*, memories can interfere with the retrieval of other memories)(Atkinson and Shiffrin, 1977) or poor encoding (*i.e.*, memories cannot be accessed because it was poorly encoded, or the concept was not embedded in a schema).

People solve problems in the working memory by comparing information from their environment with information retrieved from their long-term memory. Four categories were identified as a potential barrier to solving a problem: (1) inability to recall, (2) inability to apply or understand, (3) poorly understood content, and (4) non-content-specific barriers (Ferguson and Bodner, 2008). The categories were common for learners memorizing declarative knowledge (*i.e.*, facts and data) without procedural knowledge (*i.e.*, skills and techniques) (Cohen *et al.*, 1997; Gupta and Cohen, 2002).

In this study, we examined delocalization concepts through the lens of IPT, specifically on how information is retrieved from long-term memory to use in working memory to solve a problem.

### 6.20.2 Reasoning framework

A scientific argument is used to persuade and justify a claim using evidence and reasoning (Osborne and Patterson, 2011). The *claim* is the position being argued, or the principle that is

trying to be conveyed. The *evidence* is the data that is used on which the claim is based (Toulmin, 1958). The *warrant* is the relationship between the claim and the evidence and why the evidence backs the claim.

Several frameworks exist to qualify reasoning, such as Type I and Type II reasoning (McClary and Talanquer, 2011; Talanquer, 2017), abstractness and abstraction (Weinrich and Sevian, 2017), rule-, case- and model-based reasoning (Strickland *et al.*, 2010), modes of reasoning (Sevian and Talanquer, 2014), and mental models (Johnson-Laird, 2010), among others. Students may use a variety of reasoning techniques to answer questions.

To analyze the written responses of questions requiring a justification in this study, we used the modes of reasoning framework proposed by Sevian and Talanquer (2014). This framework has been used to determine modes of reasoning in several studies (Sevian and Talanquer, 2014; Weinrich and Talanquer, 2015; Moon *et al.*, 2016; Moreira *et al.*, 2018; Bodé *et al.*, 2019), including similar studies that analyze students' written responses (Weinrich and Talanquer, 2015; Moreira *et al.*, 2018; Bodé *et al.*, 2019). The framework described four modes of reasoning, two non-causal modes (descriptive and relational) and two causal modes (linear and multi-component).

In *descriptive* reasoning, concepts are provided without including causality (Sevian and Talanquer, 2014). For example, a statement simply stating "This proton is more acidic" would be considered descriptive because the statement simply states a fact, without any relationships or why that fact is true. *Relational* reasoning involves outlining a relationship between two concepts; however, the underlying reason for that relationship is not explained (Sevian and Talanquer, 2014). For example, a response stating: "The proton is more acidic because of resonance" would be considered relational because the response outlines the relationship between those two concepts without explaining why those concepts are used. Causal reasoning addresses the reasons **why** a phenomenon occurs (Talanquer, 2010; Cooper, 2015; Weinrich and Talanquer, 2015; Weinrich and Sevian, 2017; Bodé *et al.*, 2019; Deng and Flynn, 2021) and implies a cause-and-effect relationship between components. In *linear causal* reasoning, the relationship between concepts is present and the reason is stated for why the concepts are important and how they relate to the claim (Sevian and Talanquer, 2014) . For example, consider the statement

"The proton is more acidic because resonance will better stabilize its conjugate base, and a more stable conjugate base means a stronger acid". In this statement, the reason is stated for why resonance is important—higher stability. In *multi-component causal* reasoning, multiple linear causal relationships are involved. This type of reasoning involves weighing multiple factors and explaining why each is important, often involving an analysis of why one factor is dominant.

## 6.21 Methods

### 6.21.1 Settings and course

Participants in the study were students in Organic Chemistry I or II courses at a large, research-intensive Canadian university. The University of Ottawa's Research Ethics Board approved this study as a secondary use of data (H03-15-18).

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 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 Chemistry I course has a required, associated laboratory section (3 hours biweekly). The Organic Chemistry II course has a laboratory course that runs concurrently and is only required for some programs (3 hours weekly). The organic chemistry courses use a principles and patterns of mechanisms curriculum; in that curriculum, the electron-pushing formalism is explicitly taught before deeper concepts of reactivity are addressed, reactions are taught in a gradient of difficulty and sections are organized by governing mechanism (Flynn and Ogilvie, 2015; Flynn and Featherstone, 2017; Carle *et al.*, 2020).

### 6.21.2 Questions analyzed in this study

Twelves questions were selected (Figure 6.15) from exam questions in Organic Chemistry I and II courses as they represented the intended LOs identified in previous work (Carle and Flynn, 2020). The questions were chosen based on the available exam questions that aligned with delocalization LOs (Table 1). Several of the questions assessed multiple LOs (Questions 1,2,3,8) and each LO was assessed once, at least partially. LO2 (*Draw*) was assessed in three ways: explicitly (Question 1 and 12), implicitly (Question 2), and within a mechanism (Question 3 and Question 8). Explicit questions stated that resonance was required and was not embedded within

a mechanism or context. Implicit questions are questions that do not state in the prompt that delocalization is required. Mechanistic question explicitly required delocalization within but within a mechanism.

*Table 6.1. Selected questions aligned with the LOs.*

| | Organic Chemistry I | Organic Chemistry II |
|---|---|---|
| LO1 (Identify) | Q2 (Exam 2, N = 286) | No questions available |
| LO2 (Draw) | Q1 (Exam 1, N = 288) | Q8 (Exam 3, N = 296) |
| | Q2 (Exam 2, N = 286) | Q12 (Exam 4, N = 389) |
| | Q3 (Exam 2, N = 286) | |
| LO3 (Contribution) | Q1 (Exam 1, N = 288) | No questions available |
| | Q3 (Exam 2, N = 286)[a] | |
| LO4 (Hybrid) | Q1 (Exam 1, N = 288) | Q12 (Exam 4, N = 73)[b] |
| LO5 (Hybridization) | Q1 (Exam 1, N = 288)[a] | No questions available |
| LO6 (Aromaticity) | Q5 (Exam 2, N = 286)[a] | Q7 (Exam 3, N = 296)[a] |
| LO7 (Stability) | Q4 (Exam 2, N = 286)[a] | Q6 (Exam 3, N = 296)[a] |
| LO8 (Acid/base) | Q2 (Exam 2, N = 286) | Q10 (Exam 4, N = 389)[a] |
| LO9 (E+/Nu) | No questions available | Q11 (Exam 4, N = 389)[a] |
| LO10 (Reaction) | Q3 (Exam 2, N = 286) | Q9 (Exam 3, N = 296)[a] |
| | | Q8 (Exam 3, N = 296) |

[a]No justification was required in the question. [b] The LO was not required but some students drew the resonance hybrid, therefore we assessed LO4

Organic Chemistry I

Exam 1, N = 284

Question 1

a. Draw all the resonance structures using the curved arrows to show electron movement.

b. Rank the resonance structures in order of contribution to the resonance hybrid and justify your answer.

c. Draw the resonance hybrid.

d. What is the hybridization of each of the following atoms?
i. O
ii. N

Question 2

a.Circle the most acidic proton in the molecule below.

b. Explain your answer by making a comparison of the two possible conjugate bases.

Exam 2, N = 288

Question 3

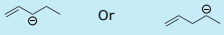
Consider the reaction of chlorobenzene with a nitronium ion (generated from $HNO_3$ and $H_2SO_4$).

a. Draw the mechanism to explain the formation of the major product, including all resonance structures.

b. Circle the most important resonance structures.

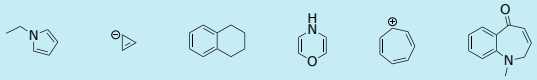
Question 4

Circle the most stable ion from the pair below.



Question 5

For each of the following species, circle the aromatic ring(s) and underline the anti-aromatic rings(s) (do nothing with the non-aromatic rings).



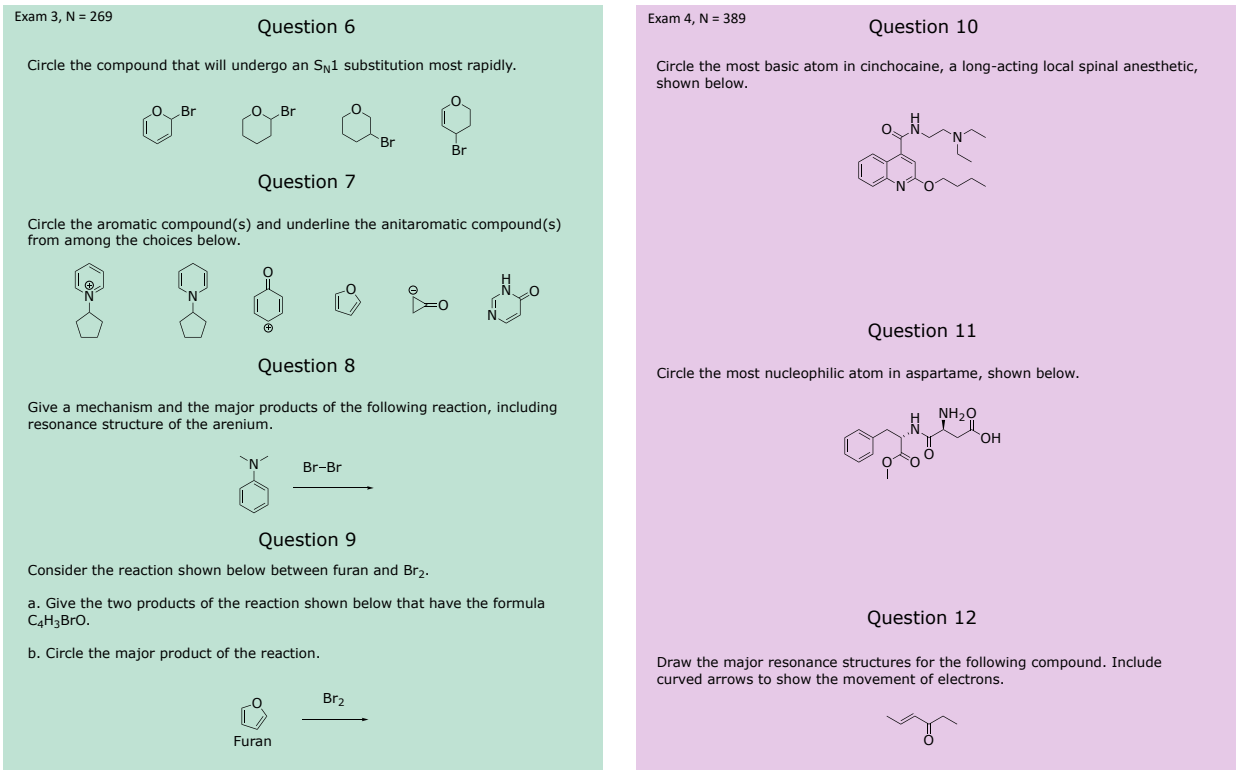



*Figure 6.15. Twelve questions assessing ten LOs. Full answers can be found in the Supporting Information.*

### 6.21.3 Coding scheme

We developed the coding scheme using inductive and deductive coding (Creswell, 2012). First, we compared students' answers to an expected answer and coded all strategies and errors encountered. After the first round of coding, the authors met and discussed the criteria required for a student to achieve a LO. The criteria came from inductive coding of the exam questions. The general criteria for LO achievement can be seen in Table 6.2, with the criteria in italics not being assessed in the question analyzed. Because many of the exam questions did not explicitly require students to justify their answers, we could not fully assess whether the full LOs have been achieved in all cases (refer to Table 6.2). If an answer contained justifications (even if not required) they were coded as fully achieved.

*Table 6.2. General criteria for LO achievement*

| LO | General criteria to achieve the LO |
|---|---|
| LO1 | Stated that delocalization can occur or drew resonance structures when not explicitly asked |
| LO2 | Correctly drew the resonance structures, with the proper bonds and charges |
| | Correctly drew the curved arrows to show electron delocalization, *i.e.*, start at the source of electrons and finish are the correct position |
| | There is no change in the sigma bonds between resonance structures |
| | There are no extra or missing structures |
| LO3 | Correctly identified the major contributor or correctly identified the order of contribution to the resonance hybrid |
| | Correctly used the evidence of charge and octet to justify their answers |
| LO4 | Correctly drew the dashed bonds representing electron delocalization |
| | Correctly drew the partial charges (not full charge) |
| LO5 | Correctly identified the hybridization of atoms in a delocalized system |
| | Justified the claim using a delocalization argument OR by drawing the resonance structures |
| LO6 | Correctly labelled the cycles as aromatic, anti-aromatic and non-aromatic |
| | Justified the claim using a delocalization argument OR by drawing the resonance structures |
| LO7 | Correctly identified the most stable structure |
| | Justified their answers by stating the cation is resonance stabilized OR by drawing the resonance structures |
| LO8 | Correctly identified the most acidic proton or basic atom |
| | Justify the answer by stating that a conjugate base can be stabilized by resonance |
| LO9 | Correctly identified the most nucleophilic atom (or electrophilic) |
| | Justify their answers by stating delocalization will lower nucleophilicity OR by drawing the resonance structures |
| LO10 | Correctly identified the regioselectivity of a reaction |
| | Justify their answers by stating delocalization will stabilize an intermediate OR by drawing the resonance structures of the intermediate |

From the inductive coding, we formed categories for the types of strategies and errors found in answers. The types of errors were grouped into three categories: (1) electron-pushing formalism (EPF), (2) structures, or (3) formalism. EPF errors involved incorrectly using the electron-pushing formalism (curved arrows) to demonstrate electron delocalization. These types of errors have been documented in previous work (Flynn and Featherstone, 2017; Carle *et al.*, 2020) and include reversed arrows, arrows from charges/atoms, or extra/missing arrows. The arrows were considered correct if they represented the correct electron delocalization; if the structures were incorrect but the arrows were correct (*i.e.*, the following resonance structure was the results of the arrows), the arrows were labelled as correct. *Delocalization formalism* errors are related to specific symbols to demonstrate delocalization (e.g., double-headed arrow to indicate the relationship between the structures). *Structure* errors consisted of errors in drawing the resonance structures, including drawing the incorrect bonds or charges, as well as extra or missing structures.

For this research, the resonance hybrid was assessed from the resonance structures drawn in the question. Therefore, if an answer had incorrect resonance structures but the correct resonance hybrid for the structures drawn, they would be coded as having achieved LO4.

For some questions, responses would contain extra information explaining the work or using strategies. The strategies were coded and categorized as (1) visualizing electrons, (2) listing properties, (3) expanding implicit features, and (4) listing rules. *Visualizing electrons* involves trying to determine where the electrons can delocalize by drawing curved arrows or resonance structures. The *Listing properties* strategy was coded when the answer contained information about the structure such as hybridization or p$K_a$ values. *Expanding implicit features* was reported in previous work(Flynn and Featherstone, 2017) and means the response contains parts of the molecules that are not explicitly shown such as lone pairs of electrons or hydrogen atoms. *Listing rules* is a strategy where the rules to solve a problem are written down, for example the rules of aromaticity or acidity.

Question 1 asked the student to justify their claim and so we analyzed their answer according to Toulmin's argumentation pattern (Toulmin, 1958). To fully achieve the LO on Question 1, the answer had to have the correct claim and correctly relate resonance contributor rules to back their claim using the following six pieces of evidence: the presence of atoms with full/absent octets of electrons and number of charges. While there are six pieces of evidence, not all were required for the answer. The answer could state that all atoms have a full octet in structures **A** and **C**, but that structure **B** does not have full octets, making structure B the minor contributor. Then to differentiate between structures **A** and **C**, the number of charges could be used. Therefore, the evidence that structure **B** having two charges was not necessary to assign the contribution of each contributor to the resonance hybrid.

Two of the questions also required a justification and were coded according to their mode of reasoning presented in the theoretical framework (Table 6.3. Criteria used to identify modes of reasoning in Question 1 and Question 2) (Sevian and Talanquer, 2014).

*Table 6.3. Criteria used to identify modes of reasoning in Question 1 and Question 2*

| Mode of reasoning | Description(Sevian and Talanquer, 2014) | Criteria – Question 1 | Criteria – Question 2 |
|---|---|---|---|
| Descriptive | Salient entities in a system are identified.<br>Explicit properties are described.<br>Explanation focused on surface features | Describes the octet (number of electrons on an atom) and/or charge<br>No link to contribution or hybrid<br>No reason was given why octet/charge is used as an explanation | Describes a proton as more acidic.<br>No link between acidity and other concepts<br>No explanation why |
| Relational | Salient entities in a system are identified.<br>Spatial or temporal relations between entities are identified.<br>Correlations between properties and behaviors are established but not explained or justified. | Describes the octet/charge<br>Links a resonance structure octet/charge to its contribution to the hybrid | Describes a proton as more acidic.<br>Links concepts together (including delocalization) |
| Linear causal | Salient entities in a system are identified.<br>Spatial or temporal organization of and connections between entities are identified.<br>Relevant direct interactions between entities invoked.<br>Proposed mechanisms involve linear cause-effect relationships and sequential chains of events. | Describes the octet/charge<br>Links a resonance structure octet/charge to its contribution to the hybrid<br>Explains WHY the octet/charge is relevant | Describes a proton as more acidic.<br>Links concepts together (including delocalization)<br>Explain that delocalization is used because it stabilizes on the conjugate base |
| Multi-component causal | Salient entities in a system are identified<br>Spatial or temporal organization of and connections between entities are identified.<br>Relevant interactions between entities are invoked.<br>Effects of several variables are considered and weighed. | Describes the octet and charge<br>Links a resonance structure octet and charge to its contribution to the hybrid<br>Explains WHY the octet and charge are relevant<br>Explain why one factor (octet) is more important to consider | Describes a proton as more acidic.<br>Links concepts together (including delocalization)<br>Uses resonance AND induction to explain their effects on the relative stability<br>Explain why one factor (delocalization) is more important |

Once we had developed the codebook, we re-coded the exams using a deductive approach. The coding book and expected answers can be found in the Supporting Information. Any statistical analyses were done using R; we used two tests: Chi-square to compared categorical data and *t*-tests to compare scores between two groups.

### 6.21.4 Validity and reliability

We addressed content validity by asking three experts to align the exam questions with the LOs. The alignment of the experts closely matched that of the authors, except for Questions 2, 3, and 8 assessing LO2 (*Draw).* The content experts mentioned that those questions only partially assessed the LO, but over 80% of the exams contained resonance structures therefore we included those as assessing LO2 (*Draw)*. Similarly, none of the experts aligned Question 12 with LO4 (*Hybrid)* since as written the hybrid is not necessary; however, 20% of students drew the resonance hybrid, and so we decided to investigate it for those students only. The experts labelled the questions in the same way the authors did as partially and fully assess each LO. In some cases, the experts mentioned that more LOs have been assessed. For example, all three experts indicated that Question 2 fully assessed LO7 (*Stability*) and LO8 (*Acid/base*) while the authors had only assigned the question to LO8.

To address inter-rater reliability (Lange, 2011), another researcher coded 15% of the exams, and we compared the codes. The percent agreement and Krippendorf $\alpha$ values were acceptable after the first round of coding for all questions except Questions 2, 3, 6, and 8, at >80% agreement and a Krippendorf $\alpha$ value above 0.7 (Krippendorff, 2004). For those questions, the raters met and discussed until agreement on the coding scheme was reached on all aspects (*i.e.*, LO achievement, errors, arguments, strategies, and reasoning). A different subset of exams was then coded and compared; the percent agreement and the calculated Krippendorf $\alpha$ values were acceptable. All percent agreements and Krippendorf $\alpha$ values are listed in the supporting information (SI).

## 6.22 Results and discussion

### 6.22.1 Overall LO achievement

The responses showed varying achievement rates on the assessed questions and LOs (Figure 6.16).

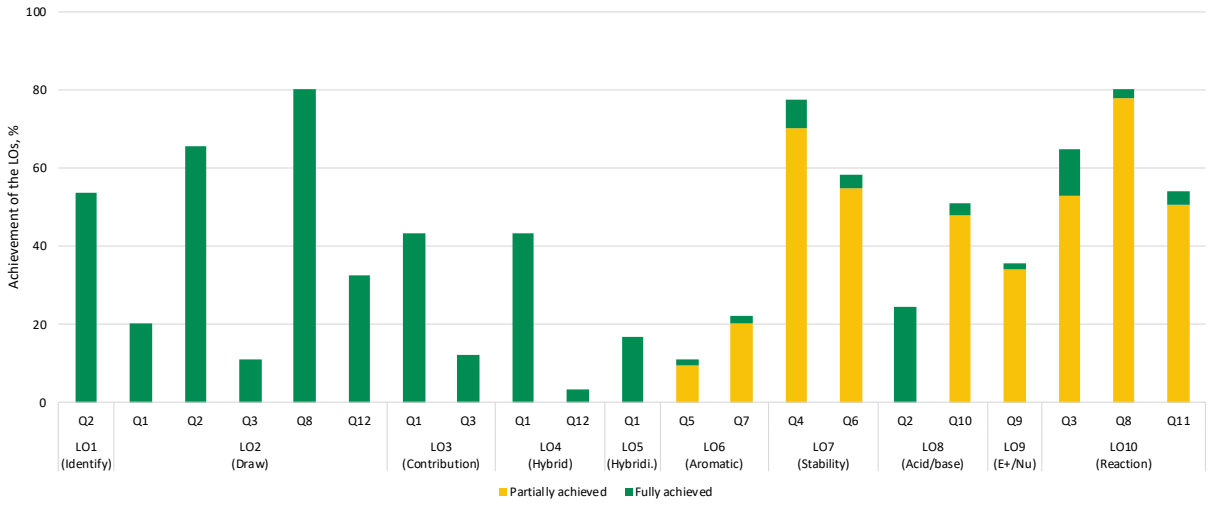



*Figure 6.16. LO achievement across all questions. Green = completely achieved, with justification, Yellow = partially achieved, had the correct answer but no justifications. Justifications were not required for the answers seen as partially achieved.*

### 6.22.2 LO1 (Identify): Some students could not or chose not to use delocalization concepts

Overall, 63% of answers to Question 2 achieved LO1 (*Identify)* (Figure 3). 54% of the answers used delocalization concepts as the justification for their claim, 15% used inductive effects, and 9% used both factors. The exam question could be answered using delocalization, inductive effects, or hybridization in the justification (argument). Analyzing either factor indicates that the enolate is more stable of the two conjugate bases. Therefore, we do not know if students had the knowledge and skills to use delocalization and simply chose to analyze the question another way.

**A few answers included delocalization, despite not being prompted.** Questions 4–8 did not explicitly have delocalization in their prompts; however, a small percentage of answers either contained resonance structures or mentioned that delocalization could occur (Question 4 = 7%, Question 5 = 8%, Question 6 = 9%, Question 7 = 10%, Question 8 = 12%) (Figure 3). Some students explicitly reasoned using delocalization to answer questions related to the other LOs, such as using delocalization to explain aromaticity, base strength, and nucleophilic sites.

Half of the answers did not use delocalization concepts for Question 2, and approximately 10% of answers explicitly used the concept in Questions 4–8. Our previous work has shown that delocalization is introduced at the beginning of the first semester of an organic chemistry course

and is not subsequently used in the course for a length of time (equivalent to a number of chapters); moreover, few practice questions represent LO1 (*Identify*)(Carle and Flynn, 2020). Therefore, students may not realize that delocalization is a factor to consider in these problems. Previous work has shown that students did not frequently consider the delocalization of electrons within a mechanism (Bhattacharyya and Bodner, 2005; Finkenstaedt-Quinn *et al.*, 2020; Petterson *et al.*, 2020). LO1 (*Identify*) is a fundamental skill for the other LOs to be achieved in later organic problems where delocalization may not be explicitly required. Therefore, being able to determine where delocalization applies is important for learners. Making the LOs explicit and providing opportunities for learners to practice this skill could help; learners likely need scaffolding (*e.g.*, explicit prompts to consider delocalization in their answers) initially.

### 6.22.3 LO2 (Draw): Varying level of achievement depending on question types, and when it was asked

LO2 (*Draw*) was addressed in five of the questions analyzed and was achieved in 13 – 65% of responses (Figure 6.17). These questions span the two-course levels, with Questions 1, 2, and 3 being from OCI, and Questions 8 and 12 being from OCII. We found that in OCI, more students achieved LO2 in an explicit question (Question 1) compared to both an implicit (Question 2) and mechanistic question (Question 3). In OCII, more students achieved LO2 on the mechanistic question (Question 8) than the explicit question (Question 12).

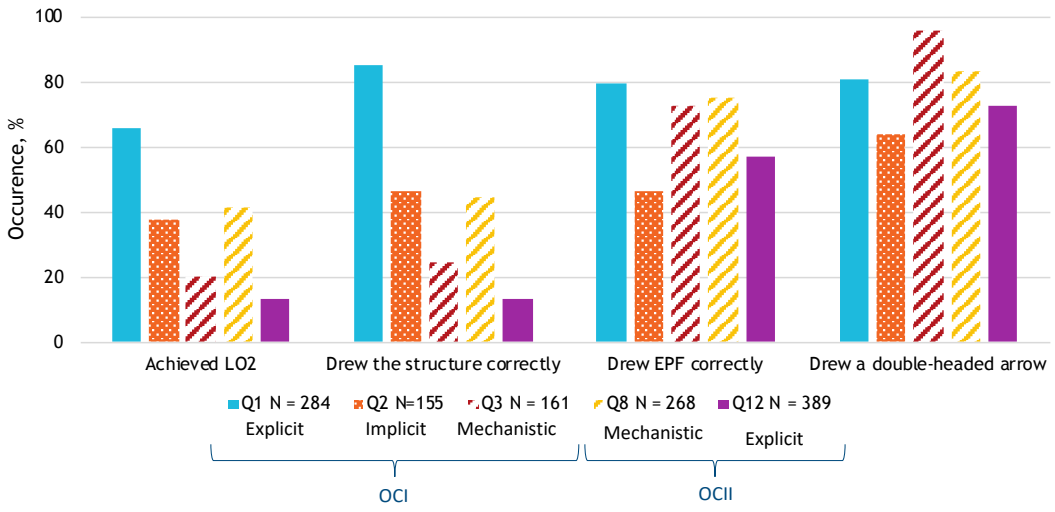



*Figure 6.17. Achievement of criteria for LO2 for each of the questions that assess LO2. Filled bars represent explicit questions, the dotted filled bar represents the implicit question, and lined-filled bars represent the mechanistic questions.*

**Answers from OCII questions contained the correct electron pushing formalism more often than OCI.** LO2 (*Draw*) was assessed in five questions and achieved to various degrees, depending on the type of question, timing of assessment, and the complexity. In OCI, most students (75%) achieved the LO in Question 1; this question explicitly asked for resonance structures and had the lowest complexity. Fewer students (40%) achieved the LO in Question 2; this question did not explicitly ask for resonance structures and the question had higher complexity as resonance concepts (or inductive effects) were needed to answer a larger question. The fewest students (18%) achieved the LO on Question 3; this question explicitly asked for resonance structures but in a later stage of the question and as part of a larger mechanism. Students may have the skills to achieved LO2 but may have difficulty using the skills when not explicitly required or in complex questions. In OCII, 32% of students correctly answered the explicit question (Question 12) but 80% of students correctly answered the mechanistic question (Question 8).

The most common errors related to the EPF were missing arrows (OCI) and incorrect arrows (OCII). In the explicit questions, there were very few answers without curved arrows, which is unsurprising since the prompt stated that curved arrows had to be drawn. The three not-explicit questions (Question 2,3,8) did not explicitly state that curved arrows were required and drawing resonance structures themselves do not require associated curved arrows. However,

most students used EPF arrows despite not being explicitly asked (80% in Question 2, 87% in Question 3, 92% in Question 8), and in over 50% of missing arrow cases, answers contained no curved arrows at all. Students may not have drawn the curved arrows because they could not or because they chose not to.

In Question 12, however, 35% of answers contained an incorrect arrow (i.e., the arrow base or point did not start/point at the right location), an error seen in less than 5% of exams for the other LO2 (*Draw*) questions. This error was often found in conjunction with a structure error since many answers had impossible structures (33% of the answers contained both structure and EPF errors). The EPF arrows with these structures were considered incorrect since the electrons could not delocalize as indicated (*i.e.*, would have created a pentavalent carbon atom).

**Many answers did not have all the expected structures.** Different educators teach delocalization differently and may have different expectations of which resonance structures should be drawn. For our analysis, we used the expected answers (Figure 6.18), which were the structures provided in the course marking scheme; these would be aligned with the instructor's expectations of the students for that specific course. Other correct structures could be included (*e.g.*, **G** and **S**).

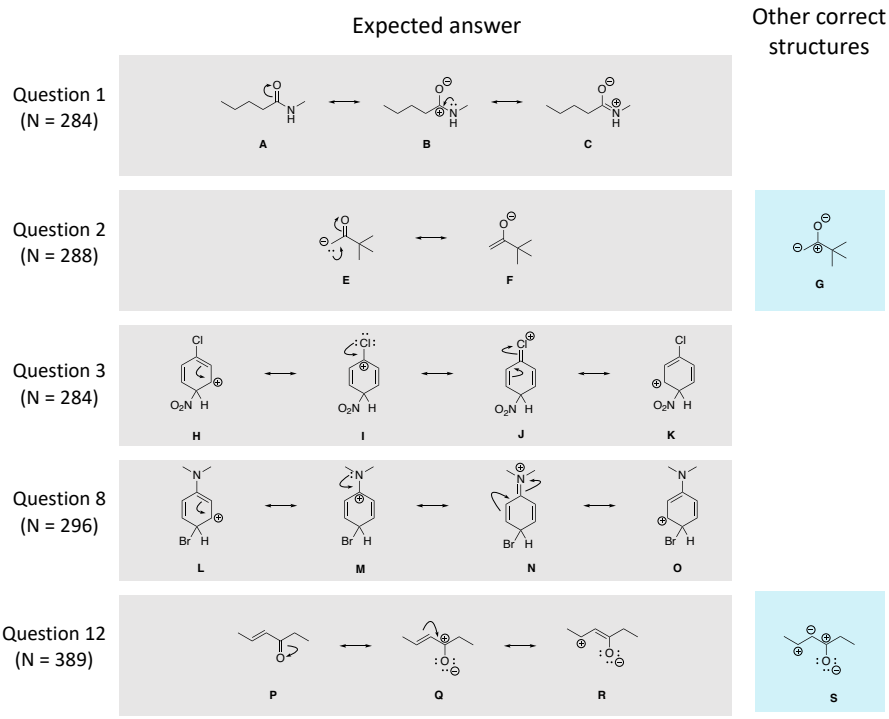



*Figure 6.18. The expected resonance structures for the five questions related to LO2 (Draw).*

For Question 1, 90% of the answers contained all three correct structures, and 2% omitted structure **B**. For the implicit Question 2, which has only two major resonance structures, 23% of students drew structure **E** as the product of the acid–base reaction; however, many did not draw **F** despite mentioning in their written answer that resonance stabilized the conjugate base. Without at least seeing a drawing of the other resonance forms, we cannot tell if students knew how resonance was involved in the molecule.

In Question 3, 60% of the answers did not include resonance structure **J**. This structure is the highest contributor to the resonance hybrid (since all atoms have a filled octet). One factor for missing this structure could be that the electrons come from outside of the ring; however, only 20% of the answers were similar in Question 8 (*i.e.*, omitting structure **N**). The different functional groups attached to the rings may have made it easier for students to cue delocalization of electrons from a nitrogen atom (with a lone pair) compared to a chlorine atom.

While answers to Question 8 showed structure **N** in their answer, 17% of answers omitted structure **L/O.** 21% of the answers drew the lone pair on the nitrogen delocalizing into the ring as part of the nucleophilic attack.

In Question 12 (OCII), structure **Q** was omitted in 78% of answers. A quarter of the answers used two curved arrows to arrive at structure **R**, bypassing structure **Q** (Figure 6.19). Structures **N** and **Q** are minor contributors to the hybrid and could be why structures were omitted. The questions did not ask for a specific number of resonance structures, so omitting a structure (especially a minor one) could be expected. Although the minor structure is important to identifying electrophilic and nucleophilic sites (LO9), it could be deemed unimportant when showing electron delocalization. For educational purposes, clear communication is needed regarding the purpose of the question (*e.g.*, draw the resonance structures that reveal they main electrophilic sites in the molecule), the specific expectations (*e.g.*, draw the four most important resonance structures for the following molecule), or other relevant information.

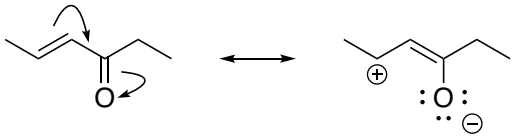


*Figure 6.19. Resonance structures for Question 12, missing two minor contributors.*

**Impossible structure errors related to the application of delocalization concepts in OCII.** Two types of structure errors were more prevalent in OCII than OCI; drawing impossible structures and drawing an incorrect reaction (Figure 6.20). Only 4% of answers in Question 1 contained these structure-related errors but 40% of exams in Question 12 contained these errors. A few answers contained a reaction (7%) in OCII and a negligible (<1%) amount in OCI. These errors were also minimal (<5%) in the mechanistic questions. The students were different in each of the courses and we have not analyzed how various sections of the courses have been taught, so we cannot make claims about the students' gains or losses in knowledge from one course to another.

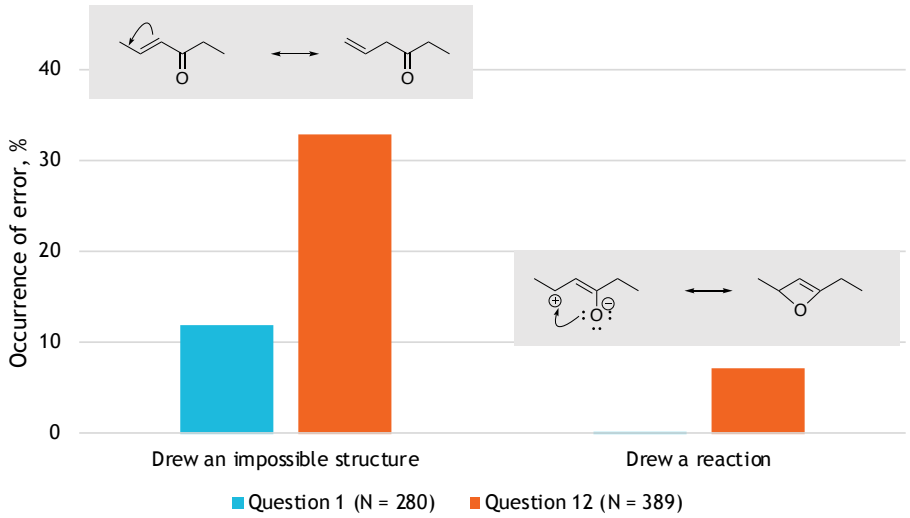



*Figure 6.20. Comparison of errors in drawing the resonance structures.*

**Most students' answers had the correct double-headed arrow** to show the relationship between resonance structures: 82% for Question 1, 85% for Question 2, 99% for Question 3, 84% for Question 8 and 72% for Question 12. The remaining answers included reaction arrows (*i.e.*, one-directional), equilibrium arrows, or nothing.

LO2 (*Draw)* achievement seemed dependent on how the resonance structures were prompted. In OCI, students were more successful on explicit questions. Drawing resonance structures is typically taught at the beginning of the organic chemistry course sequence and would therefore be expected in the OCI final exam. There is typically a gap between chapters that use and practice delocalization, which could lead to poor recall of the subject matter or poor application of the concept (Ferguson and Bodner, 2008). However, OCII students were more successful than OCI students when drawing resonance structures within a mechanism. Electrophilic aromatic substitutions are taught at the end of OCI (Flynn and Ogilvie, 2015), meaning that students in the OCI cohort had less practice using delocalization within a reaction mechanism and OCII students would have had more opportunities to practice, both using delocalization within a mechanism and using the curved arrows to draw mechanism. In Question 3 (OCI) more than half the answers did not contain resonance structures. Many of these answers had incorrect mechanisms indicating that to achieve LO2 (*Draw*) within a reaction, prior skill with the EPF and reactivity is needed.

Of the three main types of errors (structure, EPF, and delocalization formalism), structure errors were most common – especially in OCII. The delocalization formalism errors were minimal; 3–20% of answers contained a formalism error. EPF errors were also minimal for Question 1 (<20%) but were high for the other questions, especially Question 2 (64%) and Question 12 (43%). Both questions with high EPF errors also contained a high amount of structure errors (Question 2: 46%, Question 12: 13%). For Question 2, many students drew reversed arrows and had the wrong charge. In Question 12, over 35% of students drew impossible structures, with incorrect use of the EPF.

Several answers did not include all the expected structures seen in Figure 5 (Question 1: 3%, Question 2: 11%, Question 3: 62%, Question 8: 28%, Question 12: 77%). Omitting minor resonance structures may be appropriate, depending on the context (*e.g.*, need for solving a given problem) and expectations (*e.g.*, course expectations), which need to be clearly communicated.

Impossible structures were seen in the OCII explicit question (Question 12) (Betancourt-Pérez *et al.*, 2010; Xue and Stains, 2020). This error, along with the reaction error, may indicate that students struggled to answer the questions and that they may not have recalled information about delocalization or known to use their knowledge. Previous work identified a gap in practiced questions related to delocalization spanning several (10–14) chapters (Carle and Flynn, 2020). This lack of practice questions may mean students lacked the practice to gain the skills how to answer the questions.

Most students in OCI (90%) used the correct delocalization formalism (i.e., double headed arrow), with fewer in OCII (78%). In Question 12, 20% of students used the equilibrium arrows, similar to previous work (Taber, 2002; Kim *et al.*, 2019; Xue and Stains, 2020). Using the correct arrow may not indicate that students conceptualize what the structures represent but simply that it is used to denote resonance structures. One of the main representational issues reported in the literature is the alternate conception that resonance structures are alternating or in equilibrium (Taber, 2002; Kim *et al.*, 2019; Finkenstaedt-Quinn *et al.*, 2020; Petterson *et al.*, 2020; Xue and Stains, 2020). The exam analysis could not show any indication on whether the students conceptualized that resonance structures are not real or alternating.

In most questions, the EPF was used correctly: Question 1 (80%), Question 2 (83%), Question 3 (>99%) and Question 8 (81%). Question 12 (48%) contained more incorrect arrows than the other questions, which was linked with the impossible structure error.

### 6.22.4 LO3 (Contribution): Students usually gave the correct resonance contribution order, but often lacked the justification

LO3 (*Contribution)* was fully achieved in Question 1 by 44% of the students and partially achieved in Question 3 by 10% (Figure 3); both questions came from OCI exams. Question 3 did not require a justification, therefore only the students' circling of the major contributor could be assessed.

**Most justifications contained the essential evidence to reason about their claim.** Two other pieces of evidence were found in the exams: 5% of answers mentioned the stability of the resonance structure, and 10% mentioned the electronegativity of the oxygen atom.

Six pieces of information could be leveraged as evidence to support the claim; atoms' octets on all three structures and the charge on all three structures; 21% of answers contained all six pieces of evidence. However, not all the evidence was required to justify the correct claim and 46% of answers contained enough evidence to back up their claim. Choosing to use some of the evidence shows that students were able to identify the key information required to solve the problem.

Most answers (88%) had a descriptive type of reasoning in which the answer just described the resonance structures without backing (Figure 6.21). A few answers (8%) had relational reasoning in which the answers explicitly said that the charges and octet made a resonance structure the major/minor contributor. Only three answers were causal; these three answers mentioned that charges and that lacking octets destabilized molecules. The causal answers also outlined a relationship between the stability of resonance structure and their contribution to the hybrid. These results differ from previous literature that showed mostly relational/linear reasoning (Bodé *et al.*, 2019; Deng and Flynn, 2021). However, the predominantly descriptive response format matched how the topic was taught in the OCI course; therefore, providing a descriptive argument was appropriate and expected in this case.

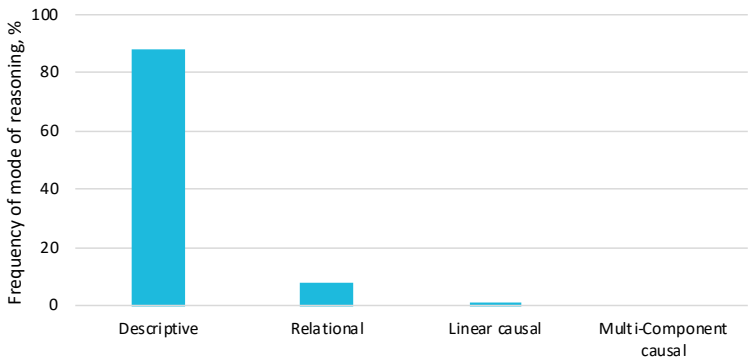



*Figure 6.21. Modes of reasoning identified in answers to Question 1 (N = 280).*

**Success on LO3 could be related to success on LO2 (*Draw*), as** only 21 students demonstrated achieving the LO in Question 3. The low success rate was related to LO2 (*Draw*), since many answers did not contain all the resonance structures, with 60% of the answers not containing the major resonance contributor. In the exams that contained structure **I**, 55% selected the correct major contributor.

LO3 (*Contribution*) was well achieved; however, the most common error in Question 3 (45% of answers that contained the major structure) involved stating that a structure containing atoms without a full octet of electrons would be the major contributor, in line with previous work (Betancourt-Pérez *et al.*, 2010). The achievement of the LO is also dependent on LO2 (*Draw)* since in Question 3 many answers did not contain the major contributor, and therefore could not select as the major contributor.

We wondered whether students would apply the concept of stability or contribution to the resonance hybrid in their answers. In this study, only 5% of the answers included "stability" when reasoning about contribution to the hybrid. Stability (a physical property) cannot be assigned to a structure that does not exist; instead, the expression "contribution to the resonance hybrid" can be used (Petterson *et al.*, 2020). Using the word stability may contribute to the alternate conception that structures are real or alternating (Petterson *et al.*, 2020). Students may still have used stability or believe stability is the key factor, without writing it down. Because students primarily used descriptive reasoning and listed the octet and charges, we do not know their mental models of delocalization. A prompt explicitly asking for causal reasoning or that asks

students to describe their mental model of resonance structures could provide more information on students' mental models.

### 6.22.5 LO4 (Hybrid): Half of the answers showed the correct hybrid structure

51% of the students successfully achieved LO4 for Question 1 (Figure 3). Question 12 did not explicitly ask for the students to draw the resonance hybrid but 19% of students included the hybrid in their answer. However, less than 50% of those answers had a correct resonance hybrid. We identified three categories of errors, related to the structure. The errors were categorized as bonds, partial charges, and formalism errors.

**There were few errors drawing dashed bonds** (<16%). Bond drawing errors were the least common for both Question 1 and 12; 90% of responses had the bonds drawn correctly on the resonance hybrid structure, and 84% for Question 12.

**The most common error was in drawing partial charges** (49%), in which students either primarily drew full charges or incorrect partial charges. The main error was a missing partial charge (Question 1 = 16%, Question 12 = 32%) or an incorrect partial charge (*i.e.*, a positive charge instead of a negative charge) (Question 1 = 12%, Question 12 = 6%). The missing partial charge error was only coded if the answer contained the resonance structure that had the charge on it. These errors would arise from mapping the information of the resonance structures onto the hybrid. While most students successfully mapped the dashed bonds, the charges contained more errors.

**The most common formalism error was labelling the hybrid as a transition state** (Question 1 = 12% and Question 12 = 16%) was the hybrid being labelled as a transition state.

**Achieving LO2 greatly increased the likelihood of achieving LO4.** Achieving LO4 (*Hybrid)* was correlated with the success of LO2 (*Draw),* $\chi^2$ (1, N = 280) = 31.55, $p < 0.001$, $\phi = 0.451$; 97% of the students who achieved LO4 also achieved LO2. In contrast, only 2% achieved LO4 without also achieving LO2 (Figure 6.22). We omitted the formalism errors for both LOs (LO2: double-headed arrow, LO4: symbol outside the bracket) because the formalism error of either would not affect the translation between representations.


*Figure 6.22. Connection between achievement of LO2 and LO4 in Question 1.*

LO4 (*Hybrid*) was achieved by 51% of students in Question 1. The most common error was a partial charge error, previous work reported that the dashed bonds errors were the most prevalent error (Betancourt-Pérez *et al.*, 2010) or that students would draw the major/minor contributor (53%) (Xue and Stains, 2020). A link between instruction and skills drawing the hybrid has been reported (Kim *et al.*, 2019; Xue and Stains, 2020), and our context may have contributed to the results obtained. The variance between previous work and our analysis could be related to the alignment on how the LOs have been intended and enacted. When teaching delocalization, a focus on the representation itself, explicitly demonstrating what information the representation provides has been shown to help students with the skills to use delocalization (Kim *et al.*, 2019; Xue and Stains, 2020).

Many students (15%) drew the correct hybrid (*i.e.*, bonds and partial charges), yet used the wrong formalism. The formalism of the hybrid is important to communicate what the structure represents; however, it does not represent whether students conceptualize the resonance hybrid. Formalism will help in communications and as such is important to teach but will not affect students understanding of the hybrid.

### 6.22.6 LO5 (Hybridization): Answer had the correct hybridization for the oxygen atom was but not the nitrogen atom

Question 1, part d asked students to label the hybridization of the oxygen and nitrogen atoms of an amide. Only 16% of students successfully labelled hybridization of both atoms (Figure 3). The oxygen was correctly identified as $sp^2$-hybridized by 64% of the students, while only 31%

correctly identified the nitrogen amide as $sp^2$-hybridized (Figure 10). The nitrogen atom was most labelled as $sp^3$-hybridized (57%). Answers of "between $sp^2$- and $sp^3$-hybridized" would have been accepted for either atom, although no students provided that answer.

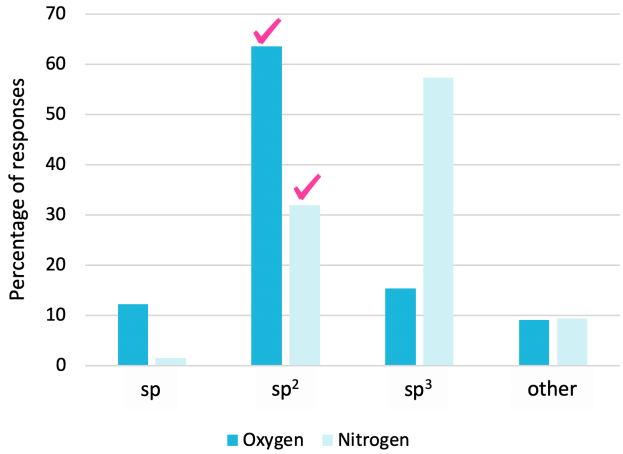



*Figure 6.23. Hybridization labels for Question 1d (N = 284). The "other" category represents answers that are not represented by the three other levels.*

LO5 (*Hybridization*) was achieved by few students. Students were more successful in assigning the hybridization of the oxygen atom than the nitrogen atom. We do not know the reasoning behind those labelling since the students were not required to justify their answers. However, the results suggest that students used the structure of the first resonance structure (major contributor) to decide on the atoms' hybridization, rather than the more accurate hybrid structure.

### 6.22.7 LO6 (Aromaticity): Answers containing a strategy (including delocalization) were more successful than not

For LO6 (*Aromaticity*), 11% of students successfully labelled all eight cycles in Question 4 (OCI), while in Question 6 (OCII), 22% successfully labelled all the cycles (Figure 3). In both questions, >50% had at least six of the eight cycles correctly identified.

Common structures, such as benzene, cyclohexane, and cyclopentane, were most often identified correctly. None of the cycles were constantly labelled wrong, with the cycle being labelled incorrectly the most was still labelled currently in 44% of the answers (Figure 6.24). Looking at the amount of correctly labelled structures we saw that most students (over 50%) correctly labelled six or more structures. In Question 5, the most common score was 6 (1 point

per correctly labelled cycle), while for Question 7 a perfect score was the most common (35% of the answers).

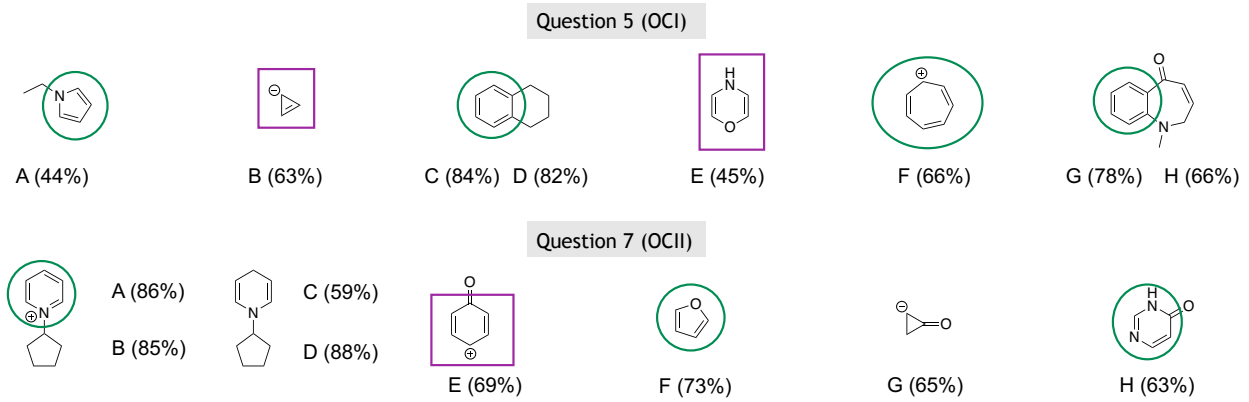



*Figure 6.24. Answers given for Q5 (OCI, N = 288) and Q7 (OCII, N = 299). Cycles that are circled in green are aromatic, cycles that are squared in purple are anti-aromatic, and the remainders are non-aromatic. The percentage below or next to each cycle is the percentage of answers with that cycle correctly labelled.*

**Answers that contained explicitly drawn electrons, listed properties, and/or expanded structures also had higher scores.** While the questions did not require justification, strategies were found in 63% of the answers in Question 5 and 78% of the answers in Question 7. For both questions, answers containing a strategy were more successful. In Question 5, the average score for an answer with a strategy was significantly higher (M = 5.7, SD = 1.68) than answers with no strategies (M = 4.5, SD = 1.80), $t(136) = 3.56$, $p = 0.0005$. Similarly for Question 7 the scores were higher for answers with a strategy (M = 6.1, SD = 1.94) than those without (M = 5.3, SD = 1.70), $t(231) = 3.14$, $p = 0.001$.

The most-used strategy for both OCI and OCII students was expanding the structure, 62% and 73% respectively. Students who used the visualizing electrons strategy, listing properties strategy and/or expanding the structure, had on average higher success in labelling the structures than students listing rules (Figure 6.25). For Question 5, listing rule by itself (M = 4.5, SD = 1.85) had no significant difference from using no strategy (M = 4.5, SD = 1.80), $t(45) = 1.82$, $p = 0.171$. The same trend was seen for Question 7, where using rules alone (M = 4.6, SD = 1.42) had no significant difference with using no strategy (M = 5.3, SD = 1.70), $t(48) = 0.91$, $p = 0.363$. But using at least one other strategy increased the average score (Question 5 = 5.27, Question 7 = 6.14).

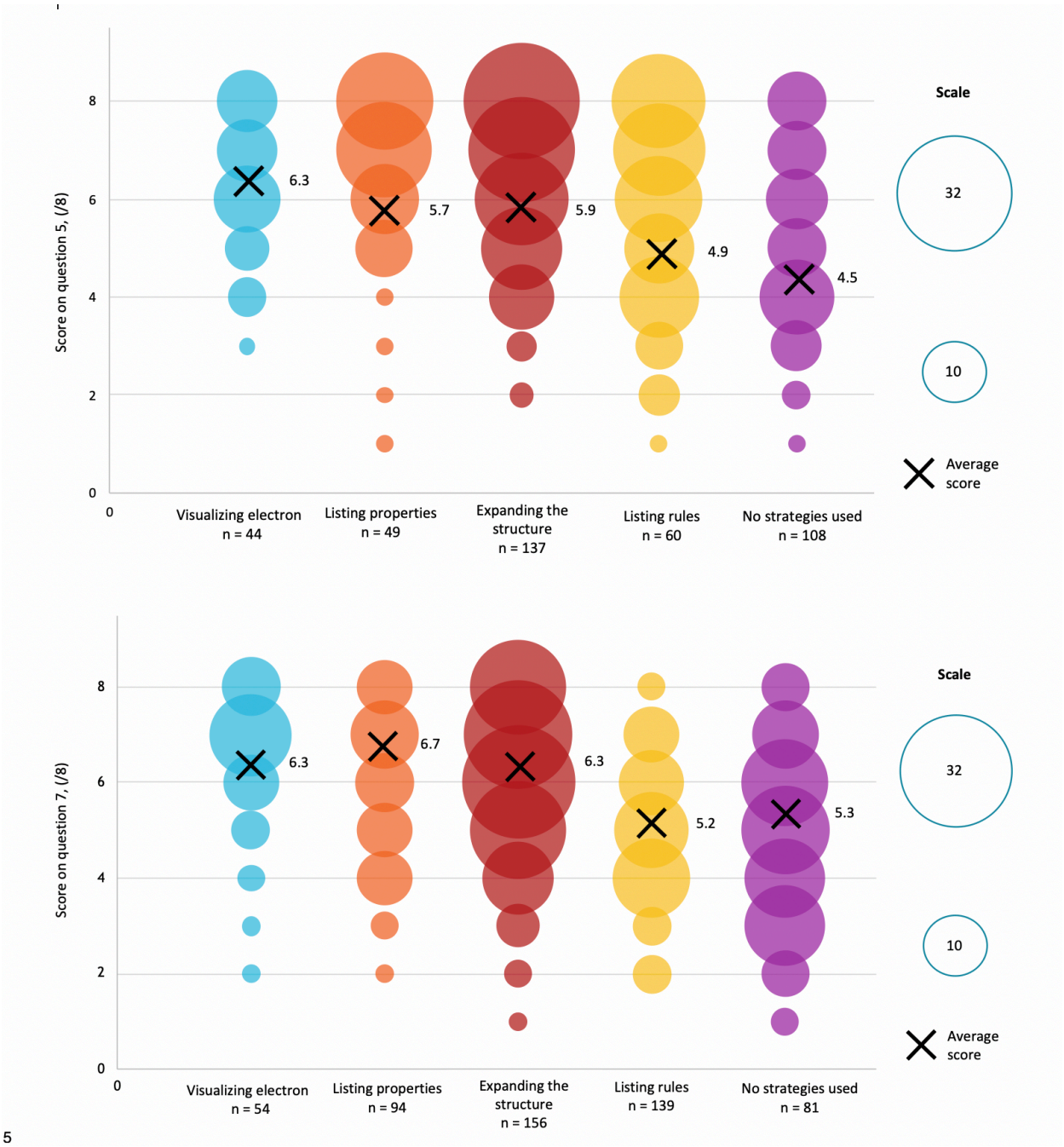



*Figure 6.25. Distribution of students' scores depending on whether they used a strategy on Question 5, N = 288 (top) and Question 7, N = 296 (bottom).*

**Strategies differed depending on the ring type**. The least number of strategies were used when determining common rings (*i.e.*, benzene, cyclohexane and cyclopentane). The low number of strategies could be due to students' familiarity with those rings since at least 82% of students labelled these correctly.

All three strategies (visualizing electrons, listing properties, and expanding the structures) were used on the three types of cycles. Listing properties (hybridization) was used more often (11%) to explain that a ring was non-aromatic (Figure 6.26). In Question 7, 79%, of answers that used the list properties strategy used it to label a $sp^3$-hybridized atom to show it was nonaromatic, while 64% of answers did so in Question 5. Students drew electrons in more often when determining if a cycle was aromatic or anti-aromatic, which was not surprising since deciding on aromaticity involves identifying the electrons involved.

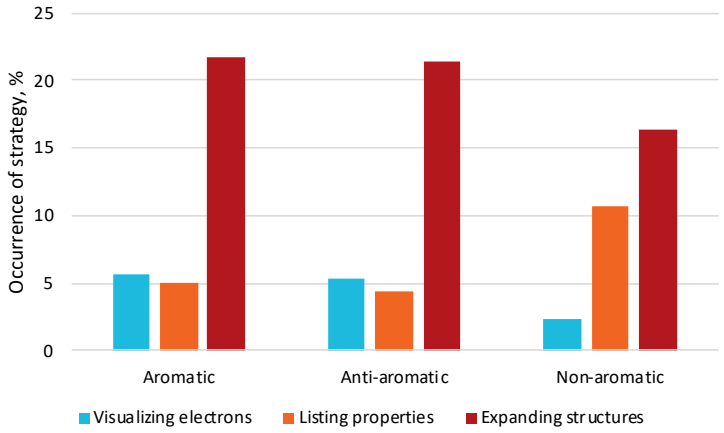



*Figure 6.26. Strategies used based on type of ring.*

LO6 (*Aromaticity*) was mostly achieved by students since they correctly labelled six of eight more than half the time. Previous work in a different context found that students seem to struggle to conceptualize aromaticity (Ealy and Hermanson, 2006), believing that any cyclic structure was aromatic (Akkuzu and Uyulgan, 2016). In the current context, the results show that students can recognize aromaticity, which may be related to the alignment of the LOs with course instructions.

Answers containing strategies of expanding the structure, visualizing electrons, and listing properties tended to have higher scores (Question 5: 5.9, Question 7: 6.4) than those that did not use those strategies (Question 5: 4.5, Question 7: 5.3). Using these strategies, students extracted additional information from the representation. Drawing the implicit information encoded in the feature can reduce cognitive load (Sweller, 1994; Paas *et al.*, 2003). Similar strategies have been connected with success in previous work (Bodé and Flynn, 2016; Flynn and Featherstone, 2017) and can be taught explicitly (Carle *et al.*, 2020). However, listing rules

showed no significant difference with using no strategies. Listing rules may result in rote memorization of defined rules, as opposed to applying concepts to extract information from the representation. Promoting and teaching the use of strategies that extract implicit information from the structures, while applying it to LO6 (*Aromaticity)* has the potential to equip learners with tools to succeed on delocalization questions.

### 6.22.8 LO7 (Stability): Using delocalization as a justification was connected with higher success

LO7 was assessed in Questions 2, 4 and 6 which was fully achieved between 7% and 14% of the time and between 59% and 78% of the time (Figure 3). Question 2 will be discussed in the next section, LO8 (*Acid/base*), since the two LOs are interconnected.

Question 4 required students to circle the most stable ion in a pair; one ion was stabilized by delocalization the other was not. 78% of students partially achieved this LO (no justification was requested); however, 14% of students added explanations or extra drawings to their answers. 7% of the students drew resonance structures or mentioned that one ion was resonance stabilized. All of the answers that invoked delocalization chose the correct ion, except one.

For Question 6, students had to choose which starting material would undergo an $S_N1$ reaction mechanism; the correct answer included the most stable carbocation that was stabilized by delocalization. Answers were correct 59% of the time. For this question, 10% of answers invoked delocalization as an extra explanation; of those, 90% identified the correct answer. The other strategies were rarely present in these answers (<5%).

LO7 (*Stability*) was well achieved; while no reasoning was required, some answers included information about delocalization. The answers that invoked delocalization (Question 4: 7%, Question 6: 10%) typically had the correct answer (Question 4: 95%, Question 6: 90%), which is aligned with previous work that showed participants seem to realize that resonance was more important in determining relative stability (Finkenstaedt-Quinn *et al.*, 2020; Petterson *et al.*, 2020). Therefore, LO1 (*Identify)* may be a building block that helps achieve LO7 (*Stability*).

### 6.22.9 LO8 (Acid/base): Answers included some causality and uses of strategies

LO8 (*Acid/base)* was achieved fully by 50% of the students (Question 2). This LO was also assessed in Question 9 but did not require a justification. In that question, 51% of students

correctly selected the most basic atom (Figure 3). Like LO7, some students demonstrated the use of delocalization to answer the questions, for those students 48% had the correct answer.

**In a multivariate acid/base question, half of the answers were correct, but it is unknown how they arrived at their claim.** Question 10 required students to select the most basic atom in a molecule, and 51% of the answers correctly selected the nitrogen atom that could not participate in delocalization (Figure 6.27). This problem is multivariate since multiple factors can affect the basicity of the atoms. Students had to determine which basic atom was least stable and to do this they must consider delocalization, hybridization, electronegativity, and inductive effects. Alternatively, they could use p$K_a$ values. Some may also have answered using heuristics or a memorized rule (*e.g.*, nitrogen is a base). Most of the students (87%) chose a nitrogen atom over an oxygen atom showing that students knew that nitrogen atoms were more basic than oxygen atoms. Without asking for explanations, we do not know the reasons for their choices.

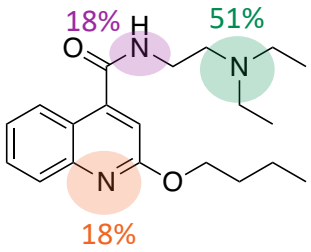



*Figure 6.27. Percentage of students who circled each site in Question 10.*

**Answer with strategies other than listing rules and properties had higher success rates.** The questions did not require justification, but many students (33%) used a strategy to answer the questions. Figure 6.28shows the percentage of students who used a specific strategy, and how many of those obtained the correct answer. As with LO6 (*Aromaticity*) answers with expanded structures and visualizing electrons were more often correct, $\chi^2$ (1, N = 389) = 32.34, $p < 0.001$, $\phi = 0.21$.

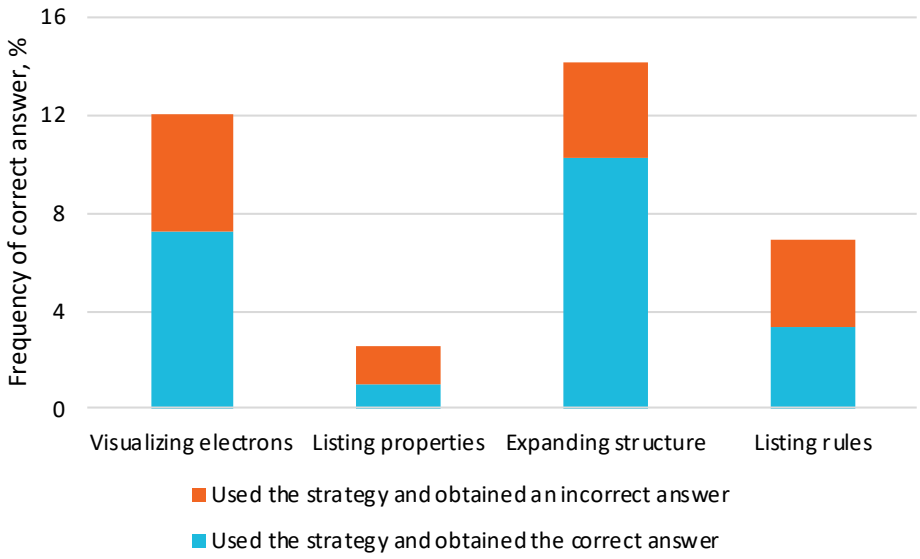



*Figure 6.28. Frequency of correct answers for Question 10 (LO8) for answers with strategies and break down per strategy.*

**The most common answer with the correct claim used delocalization as evidence.** Most answers (72%) had the correct claim—the proton they circled in part a of the question.

The evidence used for their claim could include resonance, induction, or proximity (*i.e.*, similar to induction, but only stating "close to oxygen"). 92% of the student that used resonance as evidence also had the correct claim (Figure 6.29) while 67% of the answers that used induction had the correct claim.

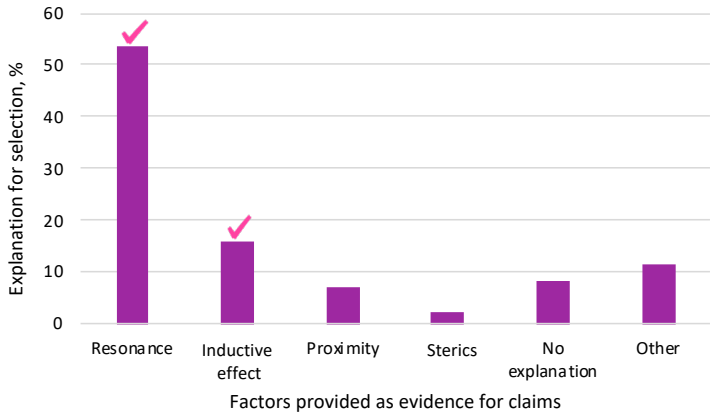



*Figure 6.29. Question 2 (acid–base): Evidence used to justify the claim (N = 280).*

The chemistry concepts used to justify the answers represent the evidence for the claims. Four concepts were expected to obtain the correct answer, and they were also the four most

commonly mentioned concepts: acid strength (69%), relative stability of the conjugate base (64%), relative base strength (54%) and one or more of resonance (55%), inductive effects (21%), and electronegativity (26%). Other concepts include charge (34%), proximity (27%), electrons (35%), and others (21%).

**Approximately half the answers had causal reasoning.** 82% of the answers showed links between concepts. Links were determined if the two concepts were connected via linking words (*e.g.*, but, therefore, as such, for example, because) or symbols (*e.g.*, =, <, >). Some answers (34%) had the appropriate links between all four concepts (*i.e.*, acid strength, conjugate base strength, conjugate base stability, and one or more of resonance, inductive effects, or electronegativity) (Figure 6.30) but others lacked one of the concept or links.

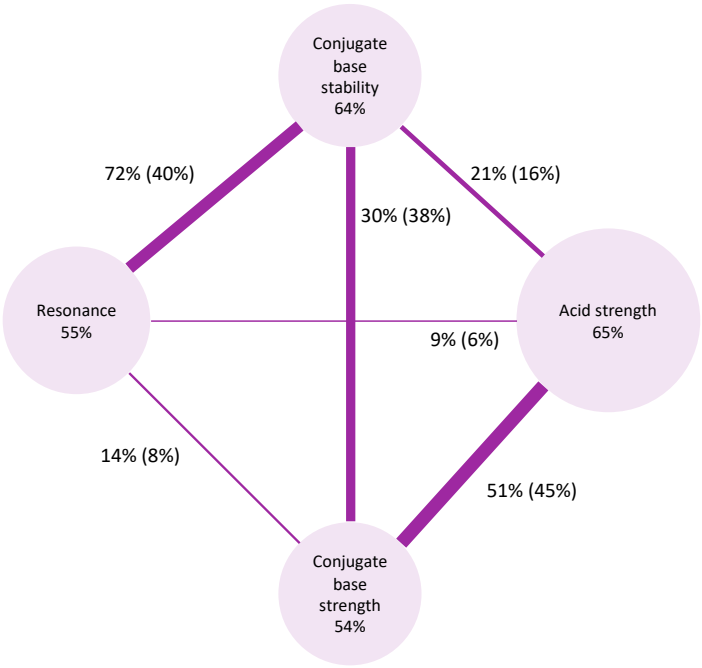



*Figure 6.30. Percentage of answers with links between concepts (n = 155). The percentage in parentheses is out of all answers, including those who did not identify resonance (N = 286).*

*Table 6.4. Examples of modes of reasoning for Question 3.*

| Mode of reasoning | Example |
|---|---|
| Descriptive | "The first conjugate base has the negative charge on the oxygen, also has resonance structures with the oxygen" Exam 227 |
| Relational | "[A] is the most acidic proton because the conjugate base is stabilized by resonance, while the other conjugate base is not" Exam 275 |
| Linear causal | "The hydrogen attached to the first carbon is the more acidic proton in the molecules, this can be seen in the conjugate base has resonance, where the alternate conjugate base does not. The molecule with the more resonance forms a more stable conjugate base, which in turns comes from a stronger acid. Therefore, the hydrogen attached to the first carbon is the more acidic proton." Exam 134 |
| Multi-component causal | "[A] is the most acidic proton because the conjugate base that would be formed if this proton was removed is more stable (weaker) compared to the C.B. [conjugate base] if the other proton was removed. The more stable C.B. would produce a stronger acid. The first C.B. is more stable since it has resonance structures that help stabilize electrons of the negatively charges carbon atom. It also has a stronger inductive effect that pull more [electron] density from the negative charge to stabilize the basic carbon." Exam 19 |

To be coded as causal, a response had to describe the relationship between an acid and its conjugate base (*e.g.*, the stronger the acid, the weaker its conjugate base) **and** explain how resonance (or induction) affects the relative stability of the conjugate bases. Examples can be seen in the linear and multi-component causal rows of

Table 6.4. The effect of resonance or induction on the conjugate base needed to be shown and the argument needed to outline why that relationship was relevant. 38% of the answers had a causal answer (linear or multi-component causal), which explained how resonance (or other factors) stabilizes one conjugate base more than the other, making the originating acid stronger (Figure 6.31).

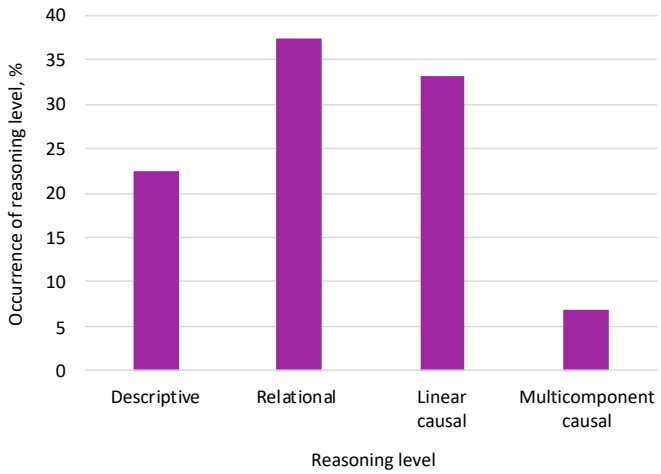


*Figure 6.31. Modes of reasoning for Question 2 (acid/base). (N = 286).*

LO8 (*Acid/base)* was achieved by half of the students in a multivariate problem that did not require reasoning, and by 34% of students when a written causal answered was expected. This LO builds on both LO1(*Identify)* and LO7 (*Stability),* and the causal link between the three LOs was seen in 38% of the answers containing delocalization.

The answers that demonstrated causal reasoning explained why resonance was used as evidence in their answers. However, many other answers did not explain why delocalization was invoked. This absence may be because the students could not or chose not to. Clearly articulating expectations to students or using reasoning scaffold(Graulich and Caspari, 2021) could promote more causal reasoning.

Question 2 mode of reasoning was similar to previous work that showed a prevalent relational/linear mode of reasoning (Bodé *et al.*, 2019; Deng and Flynn, 2021). The questions analyzed in this study and the previous study were similar as they both required students to compare two (or three) components and decide (claim) which is stronger (more likely for mechanism question) using evidence. This format has provided similar results in all three questions within our context. This work showed few (22%) descriptive answers and few (7%) multi-component causal answers, with most of the answers being relational (37%) and linear (33%).

### 6.22.10 LO9 (E+/Nu): Half the students identified the most nucleophilic atom

The LO was assessed in Question 10, required students to circle the most nucleophilic atom in a molecule, and was achieved by 53% of students (Figure 19). Students had to identify the atom that would have the highest electron density (*i.e.*, unable to delocalize its electrons) to identify the most nucleophilic atom (Figure 6.32). Most students identified that the nitrogen atom was more nucleophilic than the oxygens, however many selected the amide nitrogen. The questions did not require justification, some students (20%) used a strategy to answer the questions. We compared the frequency of correct answers between students who used the expanding structure and visualizing electrons strategies and did not use strategies (or listed ruled) and found a significant but negligible difference, $\chi^2$ (1, N = 389) = 27.01, $p < 0.001$, $\phi = 0.061$.

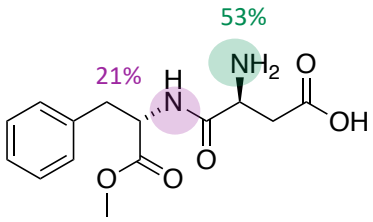



*Figure 6.32. Common answers for Question 10. The correct answer is in green.*

LO9 (*$E^+$/Nu)* was achieved by half the students and answers that explicitly contained delocalization or strategies did not have higher scores than those without. Multiple factors were involved and therefore students may have chosen not to draw or use delocalization specifically. However, multiple studies report that students often have difficulties identifying electrophiles and nucleophiles (Anzovino and Bretz, 2015; Defever *et al.*, 2015), and so we cannot say if students had difficulty with delocalization or identifying nucleophiles and electrophiles.

### 6.22.11 LO10 (Reaction): The assessments chosen did not fully assess the LO

LO10 (*Reaction*) was achieved by almost 60% of the students in Question 3 (OCI) and by 80% in Question 8 in OCII (Figure 3). Both questions asked for an electrophilic aromatic substitution (EAS) and to draw the resonance structures of the arenium ion. Question 10 also assessed LO10 regarding the bromination of a furan. This question only partially assessed the LO but was partially achieved by 54% of students.

**Electrophilic aromatic substitutions could be answered using directing groups.** The results showed that 65% of all students successfully identified the *para* or *ortho* isomer as being the major contributor to the resonance hybrid in Question 3, and 82% in Question 8. However, students may have simply memorized the rules for directing groups (*i.e.*, that chlorine is an *ortho*/*para* director) and not used delocalization in their reasoning. Most students who drew resonance structures correctly identified the *ortho* and/or *para* disubstituted benzene as being the major product (89% for Question 3, and 98% for Question 8).

**Answer that mentioned delocalization or directing groups were successful in choosing the major product in OCI.** Some of the answers (28% OCI and 30% OCII) used words to explain their claims (Figure 6.33). OCI students who used extra words to describe their answers were

successful in identifying the final major product, however students in OCII were less successful. We saw no difference between what concept was invoked (*i.e.*, resonance vs use of directing groups) in both questions.

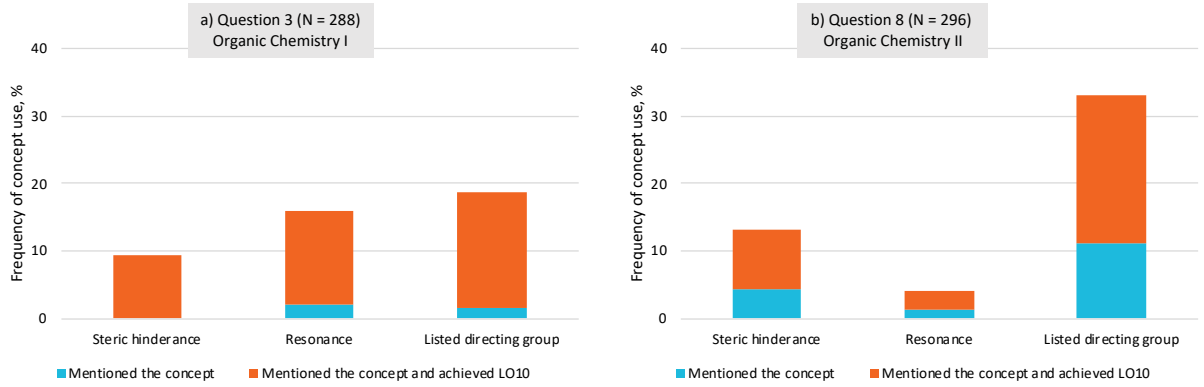



*Figure 6.33. Frequency of students mentioned a concept when not explicitly asked. (a) Question 3, OCI, N = 288, (b) Question 8, OCII, N = 296. Orange: used the concept and achieved the LO, blue: mentioned the concept but did not achieve the LO.*

Similarly, Question 9 asked students to draw the two products of the addition of bromine to furan, in which 67% of the answers had the correct products. Among those, the students were asked to identify the major product. The intermediate leading to the major product is best stabilized by delocalization, has the lowest transition state leading to its formation, and is therefore favored. Of the students who drew both isomers (67%), 53% selected the correct major product.

Although no justification was required, 13% of students either mentioned delocalization or drew resonance structures as part of their answers. There was no correlation between mentioning delocalization or drawing resonance structures and success on the question.

Students obtained the correct answer on the electrophilic aromatic substitution questions, unfortunately we cannot say if it was due to their skills using delocalization or memorizing directing groups. Therefore, more work would need to be done to identify the degree to which students are achieving this LO in different contexts. A scaffolded prompt that would explicitly require students to explain why a specific product is formed could potentially address the LO. The answers in Question 3 that contained a reference to delocalization were

mostly successful, which was also seen in a previous study (Finkenstaedt-Quinn *et al.*, 2020). However, for Question 8 half of the answers mentioning resonance were successful.

### 6.22.12 Overall key findings

Along with findings related to each LO, five other key findings emerged from our data: (1) answers had few representational errors, (2) students either did not recall information about delocalization or chose not to use it, (3) delocalization was used between 10-20% of the time when not prompted for LO6-10 (*Aromaticity, Stability, Acid/base, $E^+$/Nu, Reaction*), (4) students' reasoning aligned with course expectations, and (5) many LOs built on each other.

**Answers had few representational errors.** Few representational errors were found in the present context despite previous reports of students struggling with the representation of delocalization (Taber, 2002; Xue and Stains, 2020). The few errors about the representations itself (e.g., incorrectly drawing the dashed bonds in the resonance hybrid) were minimal and mostly related to the resonance hybrid, which differs from previous work that showed that few students identified the resonance hybrid correctly (3%) while more than half the students (53%) identified a major or minor contributor (Xue and Stains, 2020). In the current work, drawing the resonance structures was connected with higher success drawing the resonance hybrid, demonstrating the skill to translate information from one representation (resonance structures) to another (resonance hybrid) in a similar context.

Over 35% of the Question 12 answers contained at least one pentavalent carbon. This type of error could be representational since the students did not fully extract the information from the Lewis structure. In another study at the same institution, few students drew pentavalent carbons while answering questions that directly address the electron-pushing formalism and reaction mechanisms (Flynn and Featherstone, 2017; Webber and Flynn, 2018), so this error could be related to the Lewis representation of delocalization specifically.

**Strategies that helped students reason with the representations were correlated with higher achievement of the LOs.** Using strategies that involved cognitive offloading of the representation (*e.g.*, visualizing electrons, listing properties, and expanding the structure) led to the achievement of LO6 *(Aromaticity*), LO7 *(Stability*), and LO9 (*$E^+$/Nu*). These strategies involve

interacting with and using implicit information from the representation, and then using that information to solve the problem at hand. The use of those strategies indicates that students reasoned beyond surface features to understand the information decoded in the representation.

**Students either struggle to recall information about delocalization or chose not to use delocalization concepts.** In Question 2, LO1 (*Identify*) was achieved by only 54% (the question could correctly be answered in alternative ways), and delocalization was mentioned in roughly 10% of the answers in Questions 4–8. Previous work has shown that students did not frequently consider the delocalization of electrons within a mechanism (Bhattacharyya and Bodner, 2005; Finkenstaedt-Quinn *et al.*, 2020; Petterson *et al.*, 2020). The inability to recall information on delocalization would give cause for concern since without being able to identify delocalization when not explicitly told the other LOs could be impacted. Therefore, for the other LOs to be achieved in later organic problems (*e.g.*, organic synthesis, exploring mechanisms), being able to determine where delocalization applies is important. However, we do not know if this absence of delocalization information was because students chose not to mention delocalization or because they could not. Scaffolding the students to use delocalization could provide them with the skills to use delocalization in different contexts.

**Some students are using delocalization strategies when not prompted (~10%).** Delocalization was invoked by some students in all the non-explicit questions (Question 2: 54%, Question 4: 7%, Question 5: 7%, Question 6: 13%, Question 7: 10%, Question 9: 12%, Question 10: 7%, Question 11: 3%), so the prompts did cue some of the students to explicitly show delocalization. More students may have known how to use delocalization but chose not to. Answers that explicitly showed delocalization were typically successful.

Similarly, the resonance hybrid was in 19% of the answers to Question 12; the hybrid shows different information than the resonance structures and may help cue students to that information required for solving the problem. External representations have been reported to both help extract necessary information or distract from extracting the information (Paas *et al.*, 2003; DeCocq and Bhattacharyya, 2019). Therefore, drawing the hybrid may have been a sense-making device to help answer the questions by helping students reason with the representation. Alternatively, students may have drawn it in hopes of getting part-marks on a summative

assessment. In both cases, the question about delocalization cued the students to draw the hybrid.

**Students' reasoning aligned with expectations for the question and course.** Students' reasoning varied between the questions, which is most likely related to the question itself, courses expectations, and how the concepts were taught, including explanations, problem sets, and previous assessments. For Question 1, most students used a descriptive mode of reasoning and gave their answer in a tabulated/list format, while in Question 2 more answers used relational and linear causal reasoning. Previous work analyzed two comparison questions (one comparing mechanisms and the other comparing bases) similar to Question 2 and showed students typically used relational to linear causal reasoning in their responses (Bodé *et al.*, 2019; Deng and Flynn, 2021), similar to our results for Question 2. Question 1 showed predominantly descriptive reasoning, which was aligned with more rule-based reasoning where the application of rules is used to explain or draw a structure (Melanie M. Cooper *et al.*, 2010). This type of reasoning was aligned with the expectations of the course for that type of question (*i.e.*, ranking resonance structures).

**The LOs are interconnected.** Some LOs build on earlier ones. For example, students who did not identify that delocalization could occur, also could not achieve any other LO within the question. Students who achieved LO2 (*Draw)* had a higher likelihood of achieving LO4 (*Hybrid*) and LO3 (*Contribution*) and LO10 (*Reaction)*, demonstrating that the LOs are interconnected. Similarly, answers that explicitly achieved LO1 (*Identify*) were typically successful in achieving LO6 (*Aromaticity),* LO7 (*Stability*), LO8 (*Acid/base*), LO9 ($E^+$*/Nu*), and LO10 (*Reaction*). This interconnection shows that having a strong base knowledge (LO1/LO2) relates to higher success on the other LOs.

## 6.23 Conclusion

This work investigated how ten-essential delocalization LOs were achieved on exams, including strategy use and common errors. The degree of achievement of the LOs was highly varied. Errors were primarily related to drawing the resonance structures and most often related to impossible structures. Answers that contained strategies that involved drawing information from the representation (*i.e.*, drawing out electrons, listing properties, and expanding the

structures) were typically successful; however, listing rules was a strategy that was less correlated with successful answers. Two questions assessed reasoning and Question 2 (comparing relative acidity of protons) had a similar distribution of modes of reasoning as previously reported comparison questions in the literature (Bodé *et al.*, 2019; Deng and Flynn, 2021) and aligned with the course expectations and context. However, students' reasoning to explain the relative contribution of resonance structures to the resonance hybrid was primarily descriptive (87%), aligned with how this analysis type is typically taught (Carle and Flynn, 2020), the expected answer, and the course setting.

Answers contained few representational errors (*e.g.*, errors in translating between resonance structures and the hybrid). The biggest representation error was seen in Question 12, where 35% of students drew a pentavalent carbon. Since this error was minimal (<2%) in previous studies assessing EPF and mechanisms (Flynn and Featherstone, 2017; Webber and Flynn, 2018), this error could be related to the Lewis representation of delocalization specifically.

Approximately 10–20% of students used delocalization when not prompted and 19% of students used the resonance hybrid in a question that did not ask for it, potentially as a sense-making device. Similarly, when answering questions about LO6–10 that did not require justification, roughly 10% of students mentioned delocalization. In Question 2, students could use delocalization or inductive effect concepts to justify relative basicity; half of students used delocalization concepts. The other half give a variety of answers, which could indicate a preference for using another concept (inductive effects), an inability to recognize the relevance of delocalization for that context, or other reasons. Questions with explicit scaffolds or clearly stated expectations about which concepts to use could help students successfully use delocalization concepts for these questions.

Achieving LO1 (*Identify*) and LO2 (*Draw*) has led to higher success on the others LOs. In many cases, the LO1 and LO2 are part of the prior knowledge required for the latter LOs, and unsurprisingly the first two LOs would affect how students achieve the LOs.

These findings align with the context of our university and curriculum (Flynn and Ogilvie, 2015) and how the LOs have been intended and enacted (Carle and Flynn, 2020).

## 6.24 Limitations

We did not analyze in-depth how the concepts were taught. Therefore, alignment and links between student success, specific instruction, and other learning opportunities could not be assessed within the boundaries of our study. This investigation was limited by the exams and questions available, which came from a single course section. Many of the questions did not explicitly ask for justification or explanation, meaning many of the LOs could not be assessed fully and some may be overrepresented. The LOs could also be assessed in other ways.

## 6.25 Implications for research

This study analyzed students' written responses on exams, which allowed us to explore their reasoning and the chemical concepts they leveraged. Further investigation (*e.g.*, interviews) would be required to explore students' mental models of delocalization and why they used or did not use delocalization concepts in answering questions. Students who gave descriptive answers to some questions may have been able to give more sophisticated answers but did not choose to because of time constraints or believe it was not needed (*e.g.*, Question 1);(Deng and Flynn, 2021) their decisions could be explored in further research.

Students sometimes did not apply delocalization concepts when it was appropriate or expected to do so; however, we do not know why (*e.g.*, could not recall, chose not to). Students sometimes struggled in questions that required the connection of delocalization with other concepts. This relationship could be investigated further to determine why students did not identify that resonance was relevant, and how the link between delocalization and other concepts affects students' overall organic chemistry skills.

Analyzing students' responses to each LO revealed unique errors and strategies but a more in-depth analysis could provide insight into students' thought processes. A further series of questions and various prompts could provide insight into how students approach delocalization-related questions. While we outlined the strategies used, we do not know why students chose or thought to employ them. The use of these strategies could be further investigated.

## 6.26 Implications for teaching

The findings from this work could be used to inform the design and evaluation of new teaching techniques or materials. The LOs are a basis for teaching the concept of resonance that can guide instructors in their teaching and assessments. Instruction should be aligned with the desired type of reasoning (*e.g.*, relational, causal), including the taught, practiced, and assessed portions of a course or program. Some LOs were connected to being able to achieve later LOs. These links could be used to design a learning progression for the subject of delocalization, which would help align the concepts within curricula (Duncan and Hmelo-Silver, 2009).

Some students did not recall or apply information about delocalization concepts. Therefore, formative assessment or practice of the concept could benefit students. Problems in which delocalization is implicit, explicit, or mechanistic would provide students opportunities to practice the skills.

OCI students achieved LO2 more than OCII students on the explicit delocalization questions. Constantly re-enforcing the concept, and giving students practice throughout their studies could help mitigate the results. Similarly, providing formative assessments during the course sequence could help identify students' skills using delocalization and identify areas needed for review.

Answers containing certain strategies, visualizing electrons delocalization, using properties, and expanding the structures, had correct answer more than those who did not. These strategies could be scaffolded for students. Listing rules was not connected with more successful answers, and as such moving toward more causal reasoning while teaching could provide the students with the skills to reason about delocalization and use delocalization concepts.

Students have the skills to use delocalization in multiple contexts. Using more practice questions, making expectations explicit, scaffolding the concept, and promoting rezoning and strategies could provide the students the tolls to use the concept.

## 6.27 References


Akkuzu N. and Uyulgan M. A., (2016), An Epistemological Inquiry into Organic Chemistry Education: Exploration of Undergraduate Students' Conceptual Understanding of Functional Groups. *Chemistry Education Research and Practice,* **17**(1), 36–57.

Anderson J. R., (1996), ACT: A Simple Theory of Complex Cognition. *American Psychologist*, **51**(4), 355–365.

Anzovino M. E. and Lowery Bretz S., (2015), Organic Chemistry Students' Ideas About Nucleophiles and Electrophiles: The Role of Charges and Mechanisms. *Chemistry Education Research and Practice*, **16**(4), 797–810.

Arterhan C. S. C. and Schwarz B. B., (2007), The Effects of Monological and Dialogical Argumentation on Concept Learning in Evolutionary Theory. *Journal of Educational Psychology*, **99**, 626–639.

Atkinson R. C. and Shiffrin R. M., (1977), Human Memory: A Proposed System and its Control Process, in *Human Memory*, BOWER G. (ed.)., Academic Press, pp. 7–113.

Ausubel D. P., Novak J. D., and Hanesian H., (1978), *Educational psychology: A cognitive view*, Rinehart and Winston (eds.) Holt.

Berland L. K. and Reiser B. J., (2009), Making Sense of Argumentation and Explanation. *Sci. Educ.*, **93**(1), 26–55.

Betancourt-Pérez R., Olivera L. J., and Rodríguez J. E., (2010), Assessment of Organic Chemistry Students' Knowledge of Resonance-Related Structures. *Journal of Chemical Education*, **87**(5), 547–551.

Bhattacharyya G. and Bodner G. M., (2005), "It Gets Me to the Product": How Students Propose Organic Mechanisms. *Journal of Chemical Education*, **82**(9), 1402–1407.

Biggs J. and Tang C., (2011), *Teaching for Quality Learning at University*, 4th ed. The Society for Research into Higher Education.

Bodé N. E., Deng J. M., and Flynn A. B., (2019), Getting Past the Rules and to the WHY: Causal Mechanistic Arguments When Judging the Plausibility of Organic Reaction Mechanisms. *Journal of Chemical Education*, **96**(6), 1068–1082.

Bodé N. E. and Flynn A. B., (2016), Strategies of Successful Synthesis Solutions: Mapping, Mechanisms, and More. *Journal of Chemical Education*, **93**(4), 593–604.

Bodner George M. and Domin D. S., (2000), Mental Models: The Role of Representations in Problem Solving in Chemistry. *University Chemistry Education*, **4**(1), 24–30.

Brabrand C. and Dahl B., (2009), Using the SOLO taxonomy to analyze competence progression of university science curricula. *Higher education*, **58**(4), 531–549.

Brandfonbrener P. B., Watts F. M., and Shultz G. V., (2021), Organic Chemistry Students' Written Descriptions and Explanations of Resonance and Its Influence on Reactivity. *Journal of Chemical Education*, **98**(11), 3431–3441.

Carle M. S. and Flynn A. B., (2020), Essential Learning Outcomes for Delocalization (Resonance) Concepts: How are they taught, practiced, and assessed in organic chemistry? *Chemistry Education Research and Practice*, **21**(2), 622–637.

Carle M. S., Visser R., and Flynn A. B., (2020), Evaluating Students' Learning Gains, Strategies, and Errors Using OrgChem101's Module: Organic Mechanisms - Mastering the Arrows. *Chemistry Education Research and Practice*, **21**(2), 582–596.

Clayden J., Greeves N., and Warren S., (2012), *Organic Chemistry*, 2nd ed. Oxford University Press Inc.

Cohen N. J., Poldrack R. A., and Eichenbaum H., (1997), Memory for Items and Memory for Relations in the Procedural/Declarative Memory Framework. *Memory*, **5**(1–2), 131–178.

Collis K. F. and Biggs J. B., (1986), Using the SOLO Taxonomy. *Set: Research Information for Teachers*, **2**(4), 4.

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

Cooper Melanie M, Grove N., and Underwood S. M., (2010), Lost in Lewis Structures: An Investigation of Student Difficulties in Developing Representational Competence. *Journal of Chemical Education*, **87**(8), 869–874.

Cooper Melanie M., Grove N., Underwood S. M., and Klymkowsky M. W., (2010), Lost in Lewis Structures: An Investigation of Student Difficulties in Developing Representational Competence. *Journal of Chemical Education,* **87**(8), 869–874.

Creswell J. W., (2012), *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*, Pearson.

DeCocq V. and Bhattacharyya G., (2019), TMI (Too Much Information)! Effects of Given Information on Organic Chemistry Students' Approaches to Solving Mechanism Tasks. *Chemistry Education Research and Practice*, **20**(1), 213–228.

Defever R. S., Bruce H., and Bhattacharyya G., (2015), Mental Rolodexing: Senior Chemistry Majors' Understanding of Chemical and Physical Properties. *Journal of Chemical Education*, **92**(3), 415–426.

Delvigne F., (1989), A Visual Aid for Teaching the Resonance. *Journal of Chemical Education*, **66**(6), 461.

Deng J. M. and Flynn A. B., (2021), Reasoning, Granularity, and Comparisons in Students' Arguments on Two Organic Chemistry Items. *Chemistry Education Research and Practice*. **22**(3), 749-771.

Dixson D. D., Worrell F. C., Dixson D. D., and Worrell F. C., (2016), Formative and Summative Assessment in the Classroom. **5841**.

Duis J. M., (2011), Organic Chemistry Educators' Perspectives on Fundamental Concepts and Misconceptions: An Exploratory Study. *Journal of Chemical Education*, **88**(3), 346–350.

Duncan R. G. and Hmelo-Silver C. E., (2009), Learning Progressions: Aligning Curriculum, Instruction, and Assessment. *Journal of Research in Science Teaching*, **46**, 606–609.

Ealy J. B. and Hermanson J., (2006), Molecular Images in Organic Chemistry: Assessment of Understanding in Aromaticity, Symmetry, Spectroscopy, and Shielding. *Journal of Science Education and Technology*, **15**(1), 59.

Ferguson R. and Bodner G. M., (2008), Making Sense of the Arrow-Pushing Formalism Among Chemistry Majors Enrolled in Organic Chemistry. *Chemistry Education Research and Practice*, **9**(2), 102–113.

Finkenstaedt-Quinn S. A., Watts F. M., Petterson M. N., Archer S. R., Snyder-White E. P., and Shultz G. V, (2020), Exploring Student Thinking about Addition Reactions. *Journal of Chemical Education*, **97**(7), 1852–1862.

Fleming I., (2009), *Molecular Orbitals and Organic Chemical Reactions*, Student. Wiley.

Flynn A. B. and Featherstone R. B., (2017), Language of Mechanisms: Exam Analysis Reveals Students' Strengths, Strategies, and Errors when Using the Electron-Pushing Formalism (Curved Arrows) in New Reactions. *Chemistry Education Research and Practice*, **18**(1), 64–77.

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. *Journal of Chemical Education*, **92**(5), 803–810.

Gero A., (1954), Predicting Reaction of a Resonance Hybrid from Minor Canonical Structures. *Journal of Chemical Education*, **31**(3), 136.

Graulich N., (2015), The Tip of the Iceberg in Organic Chemistry Classes: How do Students Deal with the Invisible? *Chemistry Education Research and Practice*, **16**(1), 9–21.

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

Grove N. P. and Lowery Bretz S., (2012), A Continuum of Learning: From Rote Memorization to Meaningful Learning in Organic Chemistry. *Chemistry Education Research and Practice*, **13**(3), 201–208.

Günes S., Neugebauer H., and Sariciftci N. S., (2007), Conjugated Polymer-Cased Organic Solar Cells. *Chemical Reviews*, **107**(4), 1324–1338.

Gupta P. and Cohen N. J., (2002), Theoretical and Computational Analysis of Skill Learning, Repetition Priming, and Procedural Memory. *Psychological Review*, **109**(2), 401–448.

IUPAC, (2014a), Delocalization, in *IUPAC. Compendium of Chemical Terminology (the "gold book")*.

IUPAC, (2014b), Resonance, in *IUPAC. Compendium of Chemical Terminology (the "gold book")*.

Jensen W. B., (2006), More on the Nature of Resonance. *Journal of Chemical Education*, **83**(65), 223–227.

Jiang Y. and McNeill J., (2017), Light-Harvesting and Amplified Energy Transfer in Conjugated Polymer Nanoparticles. *Chemical Reviews*, **117**(2), 838–859.

Johnson-Laird P. N., (2010), Mental Models and Human Reasoning. *Proceedings of the National Academy of Sciences*, **107**(43), 18243 LP – 18250.

Kalyuga S., Ayres P., Chandler P., and Sweller J., (2003), Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist*, **38**, 23–31.

Kandlbinder P., (2014), Constructive Alignment in University Teaching. *HERDSA News*, **36**(3), 5–6.

Kerber R. C., (2006), If It's Resonance, What Is Resonating? *Journal of Chemical Education*, **83**(2), 223.

Kim T., Wright L. K., and Miller K., (2019), An Examination of Students' Perceptions of the Kekulé Resonance Representation Using a Perceptual Learning Theory Lens. *Chemistry Education Research and Practice*, **20**(4), 659–666.

Klein D., (2012), *Organic Chemistry*, John Wiley & Sons, Inc.

Kozma R. B. and Russell J., (1997), Multimedia and Understanding: Expert and Novice Responses to Different Representations of Chemical Phenomena. *Journal of Research in Science Teaching*, **34**(9), 949–968.

Kozma R. and Russell J., (2005), Multimedia Learning of Chemistry. *The Cambridge handbook of multimedia learning.*, 409–428.

Kraft A., Strickland A. M., and Bhattacharyya G., (2010), Reasonable Reasoning: Multi-Variate Problem-Solving in Organic Chemistry. *Chemistry Education Research and Practice*, **11**(4), 281–292.

Krippendorff K., (2004), Reliability in Content Analysis. *Human Communication Research*, **30**(3), 411–433.

Lange R. T., (2011), Inter-rater Reliability BT - Encyclopedia of Clinical Neuropsychology, in Kreutzer J. S., DeLuca J., and Caplan B. (eds.)., Springer New York, p. 1348.

Lin S., (2007), Aromatic Bagels: An Edible Resonance Analogy. *Journal of Chemical Education*, **84**(5), 779.

Mayer R. E. and Moreno R., (2003), Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist*, **1520**(38), 43–52.

McClary L. and Talanquer V., (2011), Heuristic Reasoning in Chemistry: Making decisions about acid strength. *International Journal of Science Education*, **33**(10), 1433–1454.

McNaught A. D. and Wilkinson A., (1997), *IUPAC. Compendium Of Chemical Terminology, 2nd ed. (the "Gold Book")*, Blackwell Scientific Publications.

Moon A., Stanford C., Cole R., and Towns M. H., (2016), The Nature of Students' Chemical Reasoning Employed in Scientific Argumentation in Physical Chemistry. *Chemistry Education Research and Practice*, **17**(2), 353–364.

Moreira P., Marzabal A., and Talanquer V., (2018), Using a Mechanistic Framework to Characterise Chemistry Students' Reasoning in Written. *Chemistry Education Research and Practice*, **20**(2), 120–131.

Mullins J. J., (2012), Hyperconjugation: A more Coherent Approach. *Journal of Chemical Education*, **89**(7), 834–836.

National Research Council, (2012), A Framework for K-12 Science Education.

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

O'Connor E., Roy K., Walsh E., Huang D., Ke D. Y. J., Campbell Brown E., et al., (2021), The Evaluation of an Integrated Growth & Goals Module to Better Equip Students with Learning Skills in Postsecondary Courses: Systematic, Scalable, and Explicit. ChemRxiv

Ogilvie W. W., Ackyord N., Browning C. S., Deslongchamps G., Lee F., and Sauer E., (2018), *Organic Chemistry Mechanistic Patterns*, 1st ed. Nelson.

Osborne J. F. and Patterson A., (2011), Scientific Argument and Explanation: A Necessary Distinction? *Science Education*, **95**(4), 627–638.

Paas F., Renkl A., and Sweller J., (2003), Cognitive Load Theory and Instructional Design: Recent Developments. *Educational Psychologist*, **38**(1), 1–4.

Petterson M. N., Watts F. M., Snyder-White E. P., Archer S. R., Shultz G. V., and Finkenstaedt-Quinn S. A., (2020), Eliciting Student Thinking About Acid–Base Reactions *via* App and Paper–Pencil Based Problem Solving. *Chemistry Education Research and Practice*, **21**(3), 878–892.

Raycroft M. A. R. and Flynn A. B., (2020), What Works? What's Missing? An Evaluation Model for Science Curricula that Analyses Learning Outcomes Through Five Lenses. *Chemistry Education Research and Practice*, **21**(4), 1110–1131.

Richards J. C., (2013), Curriculum Approaches in Language Teaching: Forward, Central, and Backward Design. *RELC Journal*, **44**(1), 5–33.

Rushton G. T., Hardy R. C., Gwaltney K. P., and Lewis S. E., (2008), Alternative Conceptions of Organic Chemistry Topics Among Fourth Year Chemistry Students. *Chemistry Education Research and Practice*, **9**(2), 122–130.

Schunk D., (2016), *Learning Theories: An Educational Perspective*, 6th ed. Pearson.

Sevian H. and Talanquer V., (2014), Rethinking Chemistry: A Learning Progression on Chemical Thinking. *Chemistry Education Research and Practice*, **15**(1), 10–23.

Silverstein T. P., (1999), The "Big Dog-Puppy Dog" Analogy for Resonance. *Journal of Chemical Education*, **76**(2), 206.

Sim J. H. and Daniel E. G. S., (2014), Representational Competence in Chemistry: A Comparison Between Students with Different Levels of Understanding of Basic Chemical Concepts and Chemical Representations. *Cogent Education*, **1**(1), 991180.

Smith D. T., Delost M. D., Qureshi H., and Njaroarson J. T., (2016), Top 200 Pharmaceutical Products by Retail Sales.

Smith J. G., (2014), *Organic Chemistry*, 4th ed. McGraw Hill.

Solomons T. W. G. and Fryhle C. B., (2000), *Chimie organique*, Traduction. Modulo.

Solomons T. W. G. and Fryhle C. B., (2011), *Organic Chemistry*, 10th ed. John Wiley & Sons, Inc.

Starkey R., (1995), Resonance Analogy Using Cartoon Characters. *Journal of Chemical Education*, **72**(6), 542.

Stoyanovich C., Gandhi A., and Flynn A. B., (2015), Acid − Base Learning Outcomes for Students in an Introductory Organic Chemistry Course. *Journal of Chemical Education*, **92**(2), 220–229.

Strickland A. M., Kraft A., and Bhattacharyya G., (2010), What Happens when Representations Fail to Represent? Graduate Students' Mental Models of Organic Chemistry Diagrams. *Chemistry Education Research and Practice*, **11**(4), 293–301.

Sweller J., (1994), *Cognitive Load Theory*, Elsevier Inc.

Sweller J., (1999), *Instructional design in technical areas*, ACER Press.

Taber K. S., (2002), Compounding Quanta: Probing the Frontiers of Students Understanding of Molecular Orbitals. *Chemistry Education Research and Practice*, **3**(2), 159–173.

Talanquer V., (2017), Concept Inventories: Predicting the Wrong Answer May Boost Performance. *Journal of Chemical Education*, **94**(12), 1805–1810.

Talanquer V., (2010), Exploring Dominant Types of Explanations Built by General Chemistry Students. *International Journal of Science Education*, **32**(18), 2393–2412.

Talanquer V. and Pollard J., (2010), Let's Teach How We Think Instead of What we Know. *Chemistry Education Research and Practice*, **11**(2), 74–83.

Toulmin S., (1958), *The Uses of Argument*, Cambridge University Press.

Truhlar D. G., (2007), The Concept of Resonance. *Journal of Chemical Education*, **84**(5), 781.

Vollhardt P. and Schore N., (2018), *Organic Chemistry: Structure and Function*, 6th ed. W.H Freeman and Company.

Wade L. G., (2010), *Organic Chemistry*, 7th ed. Pearson Prentice Hall.

Watts F. M., Schmidt-Mccormack J. A., Wilhelm C. A., Karlin A., Sattar A., Thompson B. C., et al., (2020), What Students Write About When Students Write About Mechanisms: Analysis of Features Present in Students' Written Descriptions of an Organic Reaction Mechanism. *Chemistry Education Research and Practice*, **21**(4), 1148–1172.

Webber D. M. and Flynn A. B., (2018), How Are Students Solving Familiar and Unfamiliar Organic Chemistry Mechanism Questions in a New Curriculum? *Journal of Chemical Education*, **95**(9), 1451–1467.

Weinrich M. L. and Sevian H., (2017), Capturing Students' Abstraction While Solving Organic Reaction Mechanism Problems Across a Semester. *Chemistry Education Research and Practice*, **18**(1), 169–190.

Weinrich M. L. and Talanquer V., (2016), Mapping Students' Modes of Reasoning When Thinking About Chemical Reactions Used to Make a Desired Product. *Chemistry Education Research and Practice*, **17**(2), 394–406.

Wiggins G. and McTighe J., (1998), What is Backward Design. *Understanding by design*, **1**, 7–19.

Xue D. and Stains M., (2020), Exploring Students' Understanding of Resonance and Its Relationship to Instruction. *Journal of Chemical Education*, **97**(4), 894–902.

Zamanzadeh V., Ghahramanian A., Rassouli M., Abbaszadeh A., and Alavi- H., (2015), Design and Implementation Content Validity Study: Development of an instrument for measuring Patient-Centered Communication. *Journal of Caring Sciences*, **4**(2), 165–178.

# Chapter 7. Conclusion

Many learning tools have been created to help learners understand chemistry, such as Nomenclature101.com (Flynn *et al.*, 2014), VisChem (Tasker and Dalton, 2006) and PhET (Moore *et al.*, 2016). While some of these tools have been researched for their effectiveness in teaching (Tasker and Dalton, 2008; Kelly, 2014; Bodé *et al.*, 2016; Kelly *et al.*, 2021) not all tools have evidence that they are effective. Therefore, more evidence on the effects and impacts of these tools is needed to determine how to use them for learning. My doctoral research consisted of four studies that provided evidence on three different learning tools: the OrgChem101.com "Mastering the Arrows" online module (Carle et al., 2020), a molecular dynamic visualization (manuscript *in prep)*) and constructive alignment using LOs for delocalization (Carle and Flynn, 2020: Carle, et.al., 2021). The main research question for this thesis was: *How might learning tools (i.e., molecular dynamic visualization, learning outcomes, online module) help participants build skills and mental models of chemistry (i.e., sub-microscopic and symbolic level)?*

## 7.1 Summary of findings from each study

### 7.1.1 Mental models using molecular dynamic visualization (Chapter 4)

During my doctoral research, we used Odyssey (Wavefunction, 2019), which is a simulation of the sub-microscopic level to investigate learners' mental models of the sub-microscopic level. We invited students from Organic Chemistry and Biochemistry courses to participate. Following a pre-intervention worksheet and interview, participants used the Odyssey program controlling two different simulations. The participants then filled out a post-intervention worksheet (identical to the pre-intervention worksheet) and were interviewed again.

All participants mentioned that molecules were moving through space. Eight participants showed a collision-based mental model and explained that reactions occur through a collision of molecules. Participants' use of mental models was dependent on the context. Participants would choose different mental models depending on the context or problem at hand. Some participants explicitly mentioned this choice, especially when talking about the relative number of molecules present or the representations that they used. Previous work also reported that the prompt affected the mental model that participants used (Bongers et al., 2019). However, other studies

reported that participants did not use either a dynamic mental model or a collisions mental model in some contexts (usually linked to the symbolic level) (Popova and Bretz, 2018; Atkinson et al., 2021), but did use these models in other contexts (e.g., precipitation or pressure questions) (Kottonau, 2011; Sweeder et al., 2019; Macrie-Shuck and Talanquer, 2020). In our work we found that participants do have dynamic and collision mental models but use them only in specific contexts; sometimes participants would even acknowledge that they chose one mental model over the other.

We observed that participants had conflicting mental models about randomness and that they sometimes used deterministic thinking (as opposed to the preferred probabilistic thinking). Participants in this study stated that the sub-microscopic level was random, but they also displayed deterministic rather than probabilistic mental models. The idea that reactions are the results of a collision with a specific orientation and energy and that motion is random can be difficult for learners due to a lack of opportunity to engage in those ideas (Doerr, 2000; Talanquer, 2007, 2013; Garvin-Doxas and Klymkowsky, 2008; Jolfaee et al., 2014; Yan and Talanquer, 2015). Animations often focus on successful collisions and can appear deterministic (Tasker and Dalton, 2006). Therefore, in our study, we used a simulation (i.e., portraying the molecular world as best it can be computed) to better elicit probabilistic thinking and we showed both successful and unsuccessful collisions, as well as random motion.

Cognitive dissonance occurs when new information conflicts with previously held mental models (Festinger, 1957). We observed some participants experiencing cognitive dissonance, which resulted in them dismissing the simulation, interpreting the simulation to fit their previous mental models, or using the simulation to change their mental model.

#### *7.1.1.1 Implications and future work for molecular dynamic*

Simulations can be a great tool to showcase the probabilistic nature of the sub-microscopic level. Many animations focus on successful collisions which may re-enforce the mental models that the sub-microscopic level is deterministic. Therefore, simulations can be used to provide a probabilistic visualization for students. While our study found that students had varying mental models from deterministic to probabilistic, we do not know how these mental

models were built. Future research focusing on probabilistic mental models could provide greater insight on promoting probabilistic thinking.

While participants showed motion, collision, or randomness mental models during interviews, it may not be feasible to use interviews on large scale assessments. Assessment prompts that focus on dynamic and probabilistic mental models could provide insight into students' mental models could be used in exams or informal written assessments with larger classes (e.g., general chemistry and organic chemistry) and in research studies. A study that investigated the use of written prompts to assess students' mental models of the sub-microscopic level might also compare those mental models with students' mechanistic skills and reasoning, which could elucidate how context affects what mental models are used.

### 7.1.2 EPF using OrgChem101.com Mechanistic module (Chapter 5)

EPF is a useful tool for chemists to predict how molecules will react, but learners may struggle to use EPF since they may not understand its use (Bhattacharyya and Bodner, 2005). Therefore, OrgChem101.com's "Mastering the Arrows" online module was developed by educators to teach learners how to use EPF. The module can be used as supplemental material for students, or an educator can integrate it within a course. We used a series of interventions with pre-test/post-test design to evaluate the effectiveness of the module. The worksheets were coded for correctness, strategies, and errors. Students showed learning gains using EPF after completing the module.

After completing the module, the participants in our study could more successfully use the curved arrows representing electron movement with reactions that they had never seen before. Students had learning gains using the curved arrows for electron movement and could use the arrows more effectively to predict the product of curved arrows after using the module. These results differ from other research that showed that students struggled to use the curved arrows as a predictive tool (Bhattacharyya and Bodner, 2005; Ferguson and Bodner, 2008; Bhattacharyya, 2013). However, several studies from the Flynn Research Group have shown that teaching EPF first provides students with the skills to use the curved arrows as a problem-solving tool (Bodé and Flynn, 2016; Flynn and Featherstone, 2017; Webber and Flynn, 2018). Along with

those studies, my research shows that focusing on the meaning and language of mechanisms may be beneficial to students.

#### *7.1.2.1 Implications and future work for OrgChem101.com*

In this study, the use of OrgChem101.com's "Mastering the Arrow" module led to higher scores among participants, and an increase in the mapping strategy. This study provides evidence that the module evaluated is a tool that can successfully teach learners how to use EPF. The module can be used in class, or on its own, to help learner draw the curved arrows that denote electron movement and draw the products of a mechanism.

While our study showed that participants had high learning gains, it was unclear whether those gains would last beyond the 1-hour period of the study. Nor do we have any information about how sustained use of the OrgChem101 website might affect students' skills. Therefore, future research is needed to investigate these areas for a more holistic understanding of the effect of the module.

### 7.1.3 Delocalization using the 10 LOs (Chapters 6)

Delocalization is an organic chemistry concept that is fundamental to understanding chemical reactivity, but it is unclear what students need to learn about delocalization. Therefore, we developed and evaluated ten evidenced-based LOs that can be used in Organic Chemistry courses. LOs are useful tools for educators to design activities, assessments, and learning environments to support students' skills around delocalization through backward design (Wiggins and McTighe, 1998).While several studies have outlined that student often struggle to understand delocalization (Taber, 2002; Kim et al., 2019; Xue and Stains, 2020; Brandfonbrener et al., 2021), only one study (Betancourt-Pérez et al., 2010) has proposed what LOs are needed for students to learn about delocalization.

We conducted two studies that investigated how delocalization LOs were intended, enacted, and achieved. In the first study, we analyzed seven textbooks and exams (from nine professors across Canada) to see how delocalization was represented. After the analysis and with experts' validation we synthesized 10 essential LOs for delocalization. With the 10 LOs in hand we deductively coded the textbooks and exams to see how they were enacted. In the second

study, we analyzed XX students' answers to 12 exam questions from Organic Chemistry I and II courses from 2008 to 2016.

Some of the LOs were underrepresented in the textbooks and exams. LO1 (*Identify*), LO4 (*Hybrid*), LO5 (*Hybridization*), and LO9 (*Electrophilic/Nucleophilic*) were underrepresented while only LO2 (*Draw*) and LO8 (*Acid/base*) were well-represented; questions related to the other essential LOs were infrequently asked. We found that the LOs could be grouped as teaching LOs (i.e., LOs related to teaching the concept of delocalization) and as context LOs (i.e., LOs that used delocalization within a context). The teaching LOs (i.e., LO1–5) were more present, while LOs that used delocalization in context (LO6–10) were typically under-represented.

There is also a gap between when delocalization is taught and when it is used in context, based on our textbook analysis. Teaching delocalization is usually done at the beginning of the textbooks, right after teaching Lewis structures and as a method to teach curved arrows. The next time delocalization is present is roughly 10 chapters later in the context of aromaticity and conjugation. This gap could lead students to believe that delocalization is not an important concept. Another issue is that in the 10 chapters where delocalization is absent, some concepts related to delocalization are presented, such as acid-base, carbocation stability in E1 and $S_N1$. Conceptually, delocalization is linked to these subjects but was rarely mentioned in the textbooks.

In our analysis of exam responses, we found that answers containing strategies that involved drawing information from the representation were typically successful. These strategies include, drawing implicit atoms (e.g., hydrogens), visualization electrons (e.g., drawing curved arrows or resonance structures), listing properties (e.g., hybridization). All these strategies require the students to extract information from the structure. However, listing rules was a strategy that was less correlated with successful answers.

Some students used delocalization concepts unprompted. 10–20% of students used delocalization when not prompted and 19% of students used the resonance hybrid in a question that did not ask for it. These uses of the concepts could be sense-making device or cognitive offloading.

The LOs seem to build on each other. Achieving LO1 (Identify) is often necessary to achieve any of the following LOs. LO1 and LO2 are part of the prior knowledge required for the latter LOs, and success on LO2 (*Draw*) has led to higher success on the others LOs.

#### *7.1.3.1 Implications and future research for the delocalization project*

The 10 essential LOs can be used by educators to constructively align their course for delocalization (Biggs and Tang, 2011). The LOs can also be used for building assessment and building teaching material.

Future research would include determining the connection between the LOs. We noted some instances in which LOs built on each other but determining how those LOs affect one another could help educators develop material that would allow learners to more effectively build the knowledge required to master delocalization. A longitudinal study would help clarify when students build the knowledge to master a particular LO.

Due to the format of the exams that we analyzed (i.e., constructed responses), we do not have any information regarding how students reason about delocalization. Some research exists on how students answer some questions (Xue and Stains, 2020; Brandfonbrener et al., 2021) however, research on how students reason about all the LOs is currently not available. Reasoning using delocalization could be investigated by interviewing students who use a think-aloud as they answer questions about the LOs.

## 7.2 Future directions

As with most research, my doctoral research has answered some research questions, but it has also raised more questions. Three main questions raised by my work were:

- How did learners arrive at their answers?
- What further evidence is needed for the results presented here to be generalizable?
- How can these results be disseminated to educators?

In the following sections I give an overview of how these questions arose from my research, along with an overview of how each question might be approached for further study.

### 7.2.1 How did learners arrive at their answers?

The first question can likely be answered via research that would probe learners' thinking, y exploring the processes that students employ to arrive at their answers. Providing insights about

why specific mental models are used or why certain errors or strategies occur could help educators guide learners toward expert-like thinking.

My projects started with the synthesis of LOs for the study, however, one issue with LOs is that they are focusing – not surprisingly – on outcomes, and do not investigate the path one takes to get to the answers (Estrem, 2015). Outcomes are a snapshot of a skill, but they do not account for how someone learnt a subject. As opposed to focusing solely on an outcome, a learning progression could instead be used to provide a general pathway for learners to progress. A learning progression is a proposal of how students' knowledge and sophistication evolve during instruction (Duschl et al., 2011). The goal of a learning progression is to show how students develop conceptual understanding of a subject matter and scientific practices. A learning progression starts with the student's initial ideas (lower anchor) and ends with the desired scientific skills (upper anchors). Several intermediate levels are defined as stepping-stones for the learners. Learning progressions have the benefit of providing a progression to teach as well as conceptual change for students to progress through, as opposed to simply providing a goal.

The studies in this thesis have provided insights into the type of answers that students provide for questions about delocalization and EPF, as well as what type of mental model that students use in certain contexts. However, we did not investigate how students arrived at their answers or how they built knowledge.

in the molecular dynamic visualization study (Chapter 4), we investigated what type of mental model participants used but we did not probe extensively for each mental model (i.e., only mentioned it if the participants brought up the concepts). As such, participants may have held a mental model without using it. Further interviews probing specifically for the mental models could provide insight into what mental models are held by students, and how and why they use a specific one. In our investigation of the OrgChem101.com mechanism module (Chapter 5) and the delocalization exam analysis (Chapter 7), we only had the answers that students provided , without information about their thought processes. Trying to understand how students learn could be done via qualitative semi-structured interviews, or by using a think-aloud protocol. These new studies could provide information on which models students use as they answer questions in the moment.

Similarly, there have been few studies on how learners chronologically build their knowledge around delocalization and EPF. In the delocalization project (Chapters 6 and 7), we found that some of the LOs build on each other. Therefore, learning progressions of the 10 LOs could be investigated. Creating and testing a learning progression could provide a pathway from novice to expert that learners can follow, and educators can use in their teaching.

### 7.2.2 What further evidence is needed for the findings in this thesis to be generalizable?

Evidence in CER is different than evidence in traditional chemistry research. Molecules will typically display the same pattern of behaviors; however, humans are more varied, and their behaviours are not uniform. For this reason, findings in CER cannot be generalized to all populations after a single study (Cooper, 2018; Cooper and Stowe, 2018; Holme, 2019). Also, a single study is enough to propose a pattern. In organic chemistry, a new method will not be accepted if there is only evidence that it worked on a single substrate. A substrate scope, with varying conditions, is typically required for a method to be accepted. Similarly in CER, multiple studies with convergent findings are required to be generalizable (National Research Council, 2012; Cooper and Stowe, 2018). Some standards for a consensus level of evidence in CER have been proposed (Box 8.1) (National Research Council, 2012; Cooper and Stowe, 2018).

**A Limited Level of Evidence**

Few peer-reviewed studies of limited scope with some convergence of findings or convergence with non-peer- reviewed literature or with practitioner wisdom.

**A Moderate Level of Evidence**

A well-designed study of appropriate scope that has been replicated by at least one other similar study. Often such evidence will include both quantitative and qualitative data; OR

A few large-scale studies (e.g., across multiple courses, departments, or institutions) with similar results; OR

A moderate number of smaller-scale studies (e.g., in a single course of section) with general convergence but possibly with contradictory results. If the results are contradictory, more weight must be given to studies that reflect methodological advances or a more current understanding of teaching and learning or are conducted in more modern learning environments.

**Strong Evidence**

Numerous well-designed qualitative and/or quantitative studies, with high convergence of findings.

*Box 8.1. Characterizing the strength of conclusions supported by the evidence base. Reproduced with permission from Cooper and Stowe (2018).*

Based on these standards, most of the studies in my doctoral research would be at "a limited level of evidence" or "a moderate level of evidence".

The findings from the molecular visualization study (Chapter 4) are convergent with many other studies that show visualization help student visualize the sub-microscopic level (Sanger and Greenbowe, 2000; Sanger et al., 2000; Velázquez-Marcano et al., 2004; Tasker and Dalton, 2006, 2008; Rosenthal and Sanger, 2012, 2013; Williamson et al., 2012; Kelly and Akaygun, 2016; Kelly and Hansen, 2017; Kelly et al., 2021). However, the research in molecular visualization tackles different chemical concepts (e.g., solvation, RedOx reaction), different educational concepts (e.g., mental models, learning gains), and uses different assessments techniques (e.g., interviews, worksheets). Although there is some evidence that visualizations can be useful for student learning, individual study findings may not be as generalizable.

The literature contains some convergent evidence with the delocalization study (Chapter 6 and 7) (Taber, 2002; Betancourt- Pérez et al., 2010; Kim et al., 2019; Xue and Stains, 2020; Brandfonbrener et al., 2021). Similarly, most of the findings in the OrgChem101.com Mechanism Module study (Chapter 5) were found to be convergent with other studies within our group (Bodé and Flynn, 2016; Webber and Flynn, 2018; Flynn and Featherstone, 2017; Carle et al., 2020).

However, some divergent evidence with other studies also exists (Bhattacharyya, 2014; DeCocq and Bhattacharyya, 2019). We hypothesize that the different learning environment (i.e., the curriculum) may have led to these differences, however this hypothesis would have to be tested.

The studies in this thesis have provided evidence of learning with the learning tools studied, however they lack the scope of, large-scale implementations (e.g., multiple institutions with different populations). All data for this thesis (except for the 51 assessments analyzed in Chapter 6) were obtained from undergraduate science students at the University of Ottawa. Different populations may have different results; therefore, using data from multiple institutions, with varying instructors and curricula, could provide more generalizable information on the effect of the learning environment (e.g., instructor, pedagogy) on students' learning. Another reason some of the findings are not transferable is the type of study conducted. The study in Chapter 4 was qualitative because we collected data from nine participants; therefore, the findings cannot be generalized widely to other populations or even to the chemistry students at the University of Ottawa. Two ways to provide more convergent evidence would be to: (1) designing a replication study with different populations and/or (2) conducting multiple large-scale studies that might provide convergent findings. Similarly, all the instruments used in these studies had evidence of validity and reliability. However, further studies with the instruments would provide stronger evidence of validity and reliability for those specific instruments.

### 7.2.3 How can the results of this doctoral work be mobilized to educators?

During my doctoral studies I have investigated and explored how undergraduate organic chemistry students think about the microscopic level and the symbolic level, and I evaluated three different learning tools. The next step would be knowledge mobilization of those tools. Disseminating the information could be done via journal articles and conference presentations. However, simply sharing the information is not enough for adoption (i.e., other people using the findings in their own classrooms) (Stanford et al., 2015, 2017). Creating new material from researched-based findings can be difficult and time consuming for educators. While many chemistry education researchers develop new ways to teach that are based in evidence, uptake of those tools is often not high among educators.

Educational change is often influenced by the general context of reform, a teacher's personal profile, and structural and cultural contexts (Woodbury and Gess-Newsome, 2002). Many barriers to the adoption of new educational practices often depend on what change is being suggested. Simple changes (i.e., changing problem sets) might be more readily adopted than a complete change of curriculum.

Interactive development, interactive dissemination, and constant support are tools that have been highlighted as helping propagation (Henderson et al., 2017; Stanford et al., 2017, 2017). Interactive development means that stakeholders are involved in the design of the products. More specifically, involving educators in the design phases will allow potential users to provide input on the utility of the tool, and how they would like to use it. Finally, providing on-going support to adopters and listening to their concerns can provide more sustained levels of adoption (Stanford et al., 2015).

The findings gleaned from the molecular visualization study (Chapter 4), could lead to the development of tools to assess mental models of the sub-microscopic level. Many assessments in chemistry are at the symbolic level and few questions assess how students think of the sub-microscopic level (Laverty et al., 2016; Underwood et al., 2018). Working with educators to create, use, and evaluate new assessment questions that probe at mental models of the sub-microscopic level could lead to mobilize this research.

OrgChem101.com's modules (Chapter 5) have thousands of website uses across the world, but more knowledge mobilization efforts could help reach even more learners. Interactive workshops with organic chemistry educators could enhance the project level of knowledge mobilization. Similarly, we need to continue to involve educators and students in iterations of the design and creation of new modules. Involving educators would allow stakeholder and potential adopters to inform the design toward different needs. Involving stakeholders will make sure the tools address any common issue educators may have, and help students learn.

The findings of the delocalization studies (Chapter 6 and 7), along with the literature, demonstrate that there is a misalignment with the 10 essential LOs for delocalization. Although the LOs themselves are a useful teaching tool, the propagation of a learning tool based on the

LOs, along with aligned assessments, may be more beneficial to educators. Therefore, the design and development of aligned teaching and assessment material would be the next step in this research.

## 7.3 Implications for educators

Based on my research, the following sections provide suggestions for educators.

### 7.3.1 Molecular dynamic visualization

**Use molecular dynamic visualizations to help learners build their mental models.** Based on the results from my research, and literature, dynamic visualization can help students build a mental model of the sub-microscopic level. Specifically, using a simulation in which the learner can control the simulation and shows the probabilistic nature of the submicroscopic level would be beneficial.

**Use prompt to promote probabilistic and/or dynamic answer on assessments.** Current assessment tends to focus on the representational and macroscopic level. Therefore, using prompts that require students to articulate their mental model of the submicroscopic level could give students the opportunity to use those mental models.

**Avoid deterministic languages**. Avoiding phrases that promote deterministic (or teleological) may help students gain a probabilistic mental model.

### 7.3.2 OrgChem101.com Mechanism Module

**Use the module in class or as a supplement.** Learners who used the modules had high learning gains after using it. The modules were specifically useful for students when drawing the products of a reaction and, to a lesser extent, drawing the arrows. Those two skills are important in Organic Chemistry; therefore, the module can be used by educators to help students gain the skills.

### 7.3.3 Delocalization LOs

**Focus on the conceptual and representational aspects of delocalization**. Many reports show that learners do not conceptualize delocalization in the scientifically accepted manner (Taber, 2002; Kim et al., 2019; Xue and Stains, 2020; Brandfonbrener et al., 2021).

**Provide more opportunities to identify when delocalization may occur** Identification of delocalization is a fundamental skill for the other LOs to be achieved in later organic problems where delocalization may not be explicitly required. Therefore, being able to determine where delocalization applies is important for learners. Making the LOs explicit and providing opportunities for learners to practice this skill could help; learners will likely need scaffolding (e.g., explicit prompts to consider delocalization in their answers) initially to help with their understanding.

**Scaffold context and expectations.** Several answers did not include all the expected structures (Question 1: 3%, Question 2: 11%, Question 3: 62%, Question 8: 28%, Question 12: 77%). Omitting minor resonance structures may be appropriate, depending on the context (e.g., the need for solving a given problem) and expectations (e.g., course expectations) need to be clearly communicated.

**Provide practice throughout the teaching sequence.** Students in Organic Chemistry I showed a higher mastery of LO2 (*Draw)* then those in Organic Chemistry II. Two types of structure errors were more prevalent in OCII than OCI: drawing impossible structures and drawing an incorrect reaction. While the students were different in each of the courses and therefore, we cannot make claims about the students' gains or losses in knowledge from one course to another, this may indicate that students would benefit from more consistent delocalization practice. This error, along with the reaction error, may indicate that students struggled to answer the questions and that they may not have recalled information about delocalization or known how to use their knowledge. Previous work identified a gap in practice questions related to delocalization that spanned several (10–14) chapters (Carle and Flynn, 2020).

**Use explicit instruction using delocalization when teaching in-context LOs** (LO7-10). In the curriculum, there are several areas where delocalization may be included in instruction. Providing explicit instruction for delocalization, and practice opportunities that involve delocalization would help student gain the skills to master multiple LOs. These include and are not limited to stability of leaving groups, stabilization of ions, and reactivity.

**Teach and assess successful strategies and avoid rule-based strategies.** Some answers involved strategies that led to higher success on some of the questions, specifically (1) visualizing electrons, (2) determining hybridization of atoms in potential delocalized systems, and (3) drawing all implicit atoms and electrons. These strategies require the learners to extract information that is found within the structures drawn and may be used to reduce cognitive load (i.e., amount of information in the working memory) (Sweller, 1994; Kalyuga et al., 2003). However, answers that used rules showed similar scores to the answers that did not require rules or strategies. Therefore, teach the underlying principle of delocalization may prove beneficial for learners.

**Explicitly teach and communicate expectations.** Building LOs that are specific about reasoning and explicitly modeling them to students could help students understand expectations (Deng et al., 2022). Many of the questions we evaluated did not require reasoning or justification. Specifically, all questions related to LO6, LO7, LO9 and LO10 did not require a justification. While the questions evaluated were taken from a single professor at a single university, we found that most of the assessments from our previous studies also did not require justification (Carle and Flynn, 2020).

## 7.4 Conclusions

During my PhD studies and as part of a team, I investigated the effect of three different learning and teaching tools for chemistry (Odyssey by Wave function, Organic Chemistry: Mastering the arrows online module, LOs for delocalization). Overall, the tools were beneficial in helping learners to understand the symbolic and sub-microscopic levels. I also explored how learners used the tools, what strategies and errors they made, and how they thought about chemistry. While my research provided some evidence of the value of these tools, more research is required to provide additional evidence that these tools are useful. Overall, I hope the evaluation of the learning tools and research will help the next generation of students learn, and enjoy learning, chemistry.

## 7.5 References


American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, (2014), *Standards for Educational and Psychological Testing*.

Atkinson M. B., Croisant M., and Bretz S. L., (2021), Investigating First-Year Undergraduate Chemistry Students' Reasoning with Reaction Coordinate Diagrams when Choosing Among Particulate-Level Reaction Mechanisms. *Chemistry Education Research and Practice*, **22**(1), 199–213.

Betancourt-Pérez R., Olivera L. J., and Rodríguez J. E., (2010), Assessment of Organic Chemistry Students' Knowledge of Resonance-Related Structures. *Journal of Chemical Education*, **87**(5), 547–551.

Bhattacharyya G., (2013), From Source to Sink: Mechanistic Reasoning Using the Electron-Pushing Formalism. *Journal of Chemical Education*, **90**(10), 1282–1289.

Bhattacharyya G., (2014), Trials and Tribulations: Student Approaches and Difficulties with Proposing Mechanisms Using the Electron-Pushing Formalism. *Chemistry Education Research and Practice*, **15**(4), 594–609.

Bhattacharyya G. and Bodner G. M., (2005), "It Gets Me to the Product": How Students Propose Organic Mechanisms. *Journal of Chemical Education,* **82**(9), 1402.

Bodé N. E., Caron J., and Flynn A. B., (2016), Evaluating students' learning gains and experiences from using nomenclature101.com. *Chemistry Education Research and Practice*, **17**(4), 1156–1173.

Bodé N. E. and Flynn A. B., (2016), Strategies of Successful Synthesis Solutions: Mapping, Mechanisms, and More. *Journal of Chemical Education,* **93**(4), 593–604.

Bongers A., Northoff G., and Flynn A. B., (2019), Working with Mental Models to Learn and Visualize a New Reaction Mechanism. *Chemistry Education Research and Practice.*, **20**(3), 554–569.

Brandfonbrener P. B., Watts F. M., and Shultz G. V., (2021), Organic Chemistry Students' Written Descriptions and Explanations of Resonance and Its Influence on Reactivity. *Journal of Chemical Education*, **98**(11), 3431–3441.

Bussey T. J. and Orgill M., (2019), Biochemistry Instructors' Use of Intentions for Student Learning to Evaluate and Select External Representations of Protein Translation. *Chemistry Education Research and Practice*, **20**(4), 787–803.

Carle M. S. and Flynn A. B., (2020), Essential Learning Outcomes for Delocalization (Resonance) Concepts: How are They Taught, Practiced, and Assessed in Organic Chemistry? *Chemistry Education Research and Practice*, **21**(2), 622–637.

Carle M. S., Visser R., and Flynn A. B., (2020), Evaluating Students' Learning Gains, Strategies, and Errors Using OrgChem101's Module: Organic Mechanisms—Mastering the Arrows. *Chemistry Education Research and Practice*, **21**(2), 582–596.

Cooper M. M., (2018), The Replication Crisis and Chemistry Education Research. *Journal of Chemical Education,* **95**(1), 1–2.

Cooper M. M. and Stowe R. L., (2018), Chemistry Education Research—From Personal Empiricism to Evidence, Theory, and Informed Practice. *Chemistry Reviews*, **118**(12), 6053–6087.

Crandell O. M., Lockhart M. A., and Cooper M. M., (2020), Arrows on the Page Are Not a Good Gauge: Evidence for the Importance of Causal Mechanistic Explanations about Nucleophilic Substitution in Organic Chemistry. *Journal of Chemical Education,* **97**(2), 313–327.

DeCocq V. and Bhattacharyya G., (2019), TMI (Too Much Information)! Effects of Given Information on Organic Chemistry Students' Approaches to Solving Mechanism Tasks. *Chemistry Education Research and Practice,* **20**(1), 213–228.

Deng J. M., Carle M. S., and Flynn A. B., (2022), Students' Reasoning in Chemistry Arguments and

Designing Resources Using Constructive Alignment, in *Student Reasoning in Organic Chemistry*, Advances in Chemistry Education Series. Royal Society of Chemistry. Accepted.

Doerr H. M., (2000), How Can I Find a Pattern in This Random Data? The Convergence of Multiplicative and Probabilistic Reasoning. *Journal of Mathematical Behaviours*, **18**(4), 431–454.

Duschl R., Maeng S., and Sezen A., (2011), Learning Progressions and Teaching Sequences: A Review and Analysis. *Studies in Science Education*, **42**, 123–182.

Estrem H., (2015), Threshold Concepts and Student Learning Outcomes, in *Naming What We Know: Threshold Concepts of Writing Studies*, Adler-Kassner L. and Wardle E. (eds.)., Utah State University Press, pp. 89–104.

Ferguson R. and Bodner G. M., (2008), Making Sense of the Arrow-Pushing Formalism Among Chemistry Majors Enrolled in Organic Chemistry. *Chemistry Education Research and Practice*, **9**(2), 102–113.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.

Finkenstaedt-Quinn S. A., Watts F. M., Petterson M. N., Archer S. R., Snyder-White E. P., and Shultz G. V, (2020), Exploring Student Thinking about Addition Reactions. *Journal of Chemical Education*, **97**(7), 1852–1862.

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. *Journal of Chemical Education,* **91**(11), 1855–1859.

Flynn A. B. and Featherstone R. B., (2017), Language of Mechanisms: Exam Analysis Reveals Students' Strengths, Strategies, and Errors when Using the Electron-Pushing Formalism (Curved Arrows) in New Reactions. *Chemistry Education Research and Practice*, **18**(1), 64–77.

Flynn A. B., Teaching and Learning Support Services, and Centre for e-Learning, (2012), Organic Mechanisms | Mastering the arrows. *OrgChem101.com*.

Garvin-Doxas K. and Klymkowsky M. W., (2008), Understanding Randomness and its Impact on Student Learning: Lessons Learned from Building the Biology Concept Inventory (BCI). *CBE- LSE*, **7**(2), 227–233.

Graulich N., (2015), The Tip of the Iceberg in Organic Chemistry Classes: How do Students Deal with the Invisible? *Chemistry Education Research and Practice*, **16**(1), 9–21.

Grove N. P. and Lowery Bretz S., (2012), A Continuum of Learning: From Rote Memorization to Meaningful Learning in Organic Chemistry. *Chemistry Education Research and Practice*, **13**(3), 201–208.

Henderson C., Connolly M., Dolan E. L., Finkelstein N., Franklin S., Malcom S., et al., (2017), Towards the STEM DBER Alliance: Why we Need a Discipline-Based STEM Education Research Community. *International Journal of STEM Education*, **4**(1), 14.

Holme T. A., (2019), Reproducibility, Replication, and Generalization in Research about Teaching Innovation. *Journal of Chemical Education,* **96**(11), 2359–2360.

Johnstone A. H., (1991), Why is Science Difficult to Learn? Things are Seldom What they Seem. *Journal Computer Assisted Learning*, **7**(2), 75–83.

Jolfaee S., Zazkis R., and Sinclair N., (2014), "It Is Very, Very Random Because It Doesn't Happen Very Often": Examining Learners' Discourse on Randomness, in *Probabilistic Thinking: Presenting Plural Perspectives*, Chernoff E. J. and Sriraman B. (eds.)., Springer Netherlands, pp. 397–416.

José T. J. and Williamson V. M., (2008), The Effects of a Two-Year Molecular Visualization Experience on Teachers' Attitudes, Content Knowledge, and Spatial Ability. *Journal of Chemical Education* **85**(5), 718.

Kalyuga S., Ayres P., Chandler P., and Sweller J., (2003), Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist*, **38**, 23–31.

Kelly R. and Hansen S. J. R., (2017), Exploring the Design and Use of Molecular Animations that Conflict for Understanding Chemical Reactions. *Quimica Nova* **40,** 476-481

Kelly R. M., (2014), Using Variation Theory with Metacognitive Monitoring to Develop Insights into How Students Learn from Molecular Visualizations. *Journal of Chemical Education.*, **91**(8), 1152–1161.

Kelly R. M. and Akaygun S., (2016), Insights into How Students Learn the Difference between a Weak Acid and a Strong Acid from Cartoon Tutorials Employing Visualizations. *Journal of Chemical Education,* **93**(6), 1010–1019.

Kelly R. M., Akaygun S., Hansen S. J. R., and Villalta-Cerdas A., (2017), The Effect that Comparing Molecular Animations of Varying Accuracy has on Students' Submicroscopic Explanations*. Chemistry Education Research and Practice.*, **18**(4), 582–600.

Kelly R. M., Akaygun S., Hansen S. J. R., Villalta-Cerdas A., and Adam J., (2021), Examining Learning of Atomic Level Ideas About Precipitation Reactions with a Resources Framework. *Chemistry Education Research and Practice,* **22**(4), 886–904.

Kim T., Wright L. K., and Miller K., (2019), An Examination of Students' Perceptions of the Kekulé Resonance Representation using a Perceptual Learning Theory lens. *Chemistry Education Research and Practice*, **20**(4), 659–666.

Kottonau J., (2011), An Interactive Computer Model for Improved Student Understanding of Random Particle Motion and Osmosis. *Journal of Chemical Education*, **88**(6), 772–775.

Kozma R. B. and Russell J., (1997), Multimedia and Understanding: Expert and Novice Responses to Different Representations of Chemical Phenomena. *Journal of Research in Science Teaching*, **34**(9), 949–968

Laverty, J. T., Underwood S. M., Matz R. L., Posey L. A., Carmel J. H., Caballero M. D., et al., (2016), Characterizing College Science Assessments: The Three-Dimensional Learning Assessment Protocol. *PLOS ONE*, **11**(9), e0162333.

Linenberger K. J. and Holme T. A., (2014), Results of a National Survey of Biochemistry Instructors to Determine the Prevalence and Types of Representations Used during Instruction and Assessment. *Journal of Chemical Education*, **91**(6), 800–806.

Macrie-Shuck M. and Talanquer V., (2020), Exploring Students' Explanations of Energy Transfer and Transformation. *Journal of Chemical Education*, **97**(12), 4225–4234.

Moore E. B., Mäeots M., and Smyrnaiou Z., (2016), Scaffolding for Inquiry Learning in Computer-Based Learning Environments. *New Development in Science and Technology Education.*, 87–95.

National Research Council, (2012), *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*, National Academies Press.

Petterson M. N., Watts F. M., Snyder-White E. P., Archer S. R., Shultz G. V, and Finkenstaedt-Quinn S. A., (2020), Eliciting Student Thinking About Acid–Base Reactions via App and Paper–Pencil Based Problem Solving. *Chemistry Education Research and Practice*, **21**(3), 878–892.

Popova M. and Bretz S. L., (2018), "It's Only the Major Product That We Care About in Organic Chemistry": An Analysis of Students' Annotations of Reaction Coordinate Diagrams. *Journal of Chemical Education,* **95**(7), 1086–1093.

Rosenthal D. P. and Sanger M. J., (2013), How Does Viewing One Computer Animation Affect Students' Interpretations of Another Animation Depicting the Same Oxidation–Reduction Reaction? *Chemistry Education Research and Practice*, **14**(3), 286–296.

Rosenthal D. P. and Sanger M. J., (2012), Student Misinterpretations and Misconceptions Based on Their Explanations of Two Computer Animations of Varying Complexity Depicting the Same Oxidation–Reduction Reaction. *Chemistry Education Research and Practice*, **13**(4), 471–483.

Sanger M. J. and Greenbowe T. J., (2000), Addressing Student Misconceptions Concerning Electron Flow in Aqueous Solutions with Instruction Including Computer Animations and Conceptual Change Strategies. *International Journal of Science Education*, **22**(5), 521–537.

Sanger M. J., Phelps A. J., and Fienhold J., (2000), Using a Computer Animation to Improve Students' Conceptual Understanding of a Can-Crushing Demonstration. *Journal of Chemical Education,* **77**(11), 1517.

Stanford C., Cole R., Froyd J., Friedrichsen D., Khatri R., and Henderson C., (2015), Supporting sustained adoption of education innovations: The Designing for Sustained Adoption Assessment Instrument. *International Journal of STEM Education*, **3**(1), 1.

Stanford C., Cole R., Froyd J., Henderson C., Friedrichsen D., and Khatri R., (2017), Analysis of Propagation Plans in NSF-Funded Education Development Projects. *Journal of Science Education and Technology*, **26**(4), 418–437.

Sweeder R. D., Herrington D. G., and VandenPlas J. R., (2019), Supporting Students' Conceptual Understanding of Kinetics Using Screencasts and Simulations Outside of the Classroom. *Chemistry Education Research and Practice*, **20**(4), 685–698.

Sweller J., (1994), *Cognitive Load Theory*, Elsevier Inc.

Taber K. S., (2002), Compounding Quanta: Probing the Frontiers of Students Understanding of Molecular Orbitals. *Chemistry Education Research and Practice*, **3**(2), 159–173.

Taber K. S. and García-Franco A., (2010), Learning Processes in Chemistry: Drawing Upon Cognitive Resources to Learn About the Particulate Structure of Matter. *Journal of Learning Science*, **19**(1), 99–142.

Talanquer V., (2007), Explanations and Teleology in Chemistry Education. *International Journal of Science Education*, **29**(7), 853–870.

Talanquer V., (2013), When Atoms Want. *Journal of Chemical Education,* **90**(11), 1419–1424.

Tasker R. and Dalton R., (2006), Research into Practice: Visualisation of the Molecular World Using Animations. *Chemistry Education Research and Practice*, **7**(2), 141–159.

Tasker R. and Dalton R., (2008), Visualizing the Molecular World – Design, Evaluation, and Use of Animations, in *Visualization: Theory and Practice in Science Education*, Gilbert J. K., Reiner M., and Nakhleh M. (eds.)., Springer Netherlands, pp. 103–131.

Underwood S. M., Posey L. A., Herrington D. G., Carmel J. H., and Cooper M. M., (2018), Adapting Assessment Tasks to Support Three-Dimensional Learning. *Journal of Chemical Education.*, **95**(2), 207–217.

Velázquez-Marcano A., Williamson V. M., Ashkenazi G., Tasker R., and Williamson K. C., (2004), The Use of Video Demonstrations and Particulate Animation in General Chemistry. *Journal of Science Education and Technology*, **13**(3), 315–323.

Wavefunction, (2019), *Odyssey*, Wavefunction.

Webber D. M. and Flynn A. B., (2018), How Are Students Solving Familiar and Unfamiliar Organic Chemistry Mechanism Questions in a New Curriculum? *Journal of Chemical Education*, **95**(9), 1451–1467.

Wiggins G. and McTighe J., (1998), What is Backward Design. *Understanding by design*, **1**, 7–19.
Williamson V. M., Lane S. M., Gilbreath T., Tasker R., Ashkenazi G., Williamson K. C., and

Macfarlane R. D., (2012), The Effect of Viewing Order of Macroscopic and Particulate Visualizations on Students' Particulate Explanations. *Journal of Chemical Education* **89**(8), 979–987.

Woodbury S. and Gess-Newsome J., (2002), Overcoming the Paradox of Change without Difference: A Model of Change in the Arena of Fundamental School Reform. *Educational Policy*, **16**(5), 763–782.

Xue D. and Stains M., (2020), Exploring Students' Understanding of Resonance and Its Relationship to Instruction. *Journal of Chemical Education,* **97**(4), 894–902.

Yan F. and Talanquer V., (2015), Students' Ideas about How and Why Chemical Reactions Happen: Mapping the conceptual landscape. *International Journal of Science Education*, **37**(18), 3066–3092.

# Appendix 1: "Molecular dynamic visualization" (Chapter 4) supplemental information

## A1.1 Worksheet

1. How do you visualize what occurs at the molecular level at three different timepoints for a 1:1 mixture of acetone and water (drawn below)? Please use drawings and words.

2. For the mixture of acetone and water, describe how the molecules are moving, if at all. You can use pictures, words, or both.

3. Fluorine ($F_2$) and nitrogen dioxide ($NO_2$) are in the air around us and can react together with the following equilibria. How do you visualize what occurs at the molecular level at three different timepoints? Please use drawings and words.

$$F_2 + NO_2 \leftrightharpoons NO_2F + F$$

$$NO_2 + F \rightarrow NO_2F$$

4. What would you add to or change with your drawings or explanations of your visualization of the molecular world to make it more scientifically accurate? Please draw or explain your response.

5. What are the molecular requirements for a reaction to be successful?

6. Circle how you think reactions occur at the molecular level, with 1 molecule moving completely randomly and 10 being molecules moving purposefully toward each other.

## A1.2 Themes for randomness mental models

The emerging themes about randomness are listed in Table A1.1 along with a description and a representative quote.

*Table A1.1. Themes for randomness*

| Theme | Description | Representative quote |
|---|---|---|
| Conceptualized that the microscopic level is random | Described the microscopic level as random<br>Did not try to rationalize randomness<br>Explained that there is no purpose | P9-Post: It's mostly random, just randomly, like move bounce into each other, and there's no thought process behind them, saying calling themselves over, see, hey, come over here. I'm over, over here. It's random. |
| Struggled to conceptualize randomness | Explicitly stated that they cannot conceptualize that it's random | P4-Pre: Most like probably some sort of randomness. But to be honest, I don't, I couldn't really explain it to you. I don't know if that makes sense. So I just kind of went with the 10. To because I do know, you know, they'll move around with a purpose, obviously. |
| Adding drivers to a process | Mentioned that some factors (micro, macro) would affect how random the microscopic level is | P4-Post: I still think that a lot of the movement of the molecules depends on… Mainly the forces between the molecules and the state of matter of the substance. So I don't think a scale would encompass the variety of molecules with a variety of different interaction wouldn't be encompassed by a scale. There are too many factors to do that. |
| Believing that because reactions are efficient the microscopic world cannot be random | Explain that the fact that reaction occur show that it's not random | P6-Post: It is a random motion, but they're not just it has to be kind of, there's got to be some purpose since they are still kind of colliding with each other rather than just floating around and infinite kind of space. |
| Believing molecules have a purpose | Explain that molecules act in such a way because they have a purpose<br>Mention that molecules have a pre-determined path | P2-Pre: Because like, every molecule should have like, should have, like, it can react with something. So like, reaction can occur with each molecule under this product that forms as it can be used in many ways. So there's always purpose of for the molecules even like if you don't know like all the reactions, like all the molecules and like interactions, but what was the reaction is going on? |

## A1.3 Types of learning and cognitive dissonance

*Table A1.2. Types of cognitive dissonance*

| Theme | Participant | Description | Representative quote |
|---|---|---|---|
| No cognitive dissonance – already held a random mental model | P1 | Held the same mental model pre and post interview **AND** Has a random mental model | I: What did you think of the visualization? P1-Post: Um, I liked it. It was pretty much like, what I was visualizing the before I saw it. But I guess like, I guess that's on me for like not being able to, like, communicate that. There was no big surprises in there. […] And, like, as I did have to, like, I did see like these types of visualizations in class before, so it probably like clicked, way back then. But now there's no really, there's no real surprises anymore. |
| Interpretation simulation to fit their previous mental model | P2, P4, P9 | Provide an incorrect interpretation of the simulation to fit previous mental model OR Correctly interpreted the simulation, but mentions it is missing something from previous mental model | P2-Post: Yeah, so like, in the video, like, it kind of like also, like read my mind also, like, there is also a purpose for every reaction. Because the like that for the first one, acetone and H2O when we put them together. |
| Dismiss the simulation to retain mental model | P3, P4, P7 | Held the same mental model pre and post interview **AND** Has a conflicting mental model of randomness | P4-Post: It didn't change much from before. I still think that a lot of the movement of the molecules depends on Mainly the forces between the molecules and the state of matter of the substance. So I don't think a scale would encompass the variety of molecules with a variety of different interaction wouldn't be encompassed by a scale. There are too many factors to do that. |
| Use the simulation to change mental model | P6, P8, P10 | Changed mental model from pre to post interview **AND** Post mental model is more random than pre | P6-TA: So I'm starting to realize that there that there is quite a bit of randomness to how they're moving. It's just as they're, they're, they're they're contained in this container and kind of the randomness is just there. Once they make the connection, they form that bond. But until then, they're kind of just freely floating in the air. |

# Appendix 2: "OrgChem101.com" online module (Chapter 5) supplemental information

## A2.1 Cohort I study

Two pilot (cohort I) studies were conducted to ensure that the planned study's structure would unfold as expected and adjust as needed. The first is described in a previously published article (Visser and Flynn, 2018); the second is described below.

### A2.1.1 Participants and setting of the cohort I study

Organic chemistry I students enrolled in the 2018 winter term were invited to participate in a workshop held after their regularly scheduled classes. The researchers made an announcement during a class period and the professor teaching the course posted a recruitment text on the course online page. Pizza was provided as an incentive. Workshop attendees provided informed consent to participate in the study and could elect to participate in the workshop without having their data used for the study; nine attendees consented to have their data used in the study.

The cohort I study consisted of four parts: (1) a pre-test, (2) time allotted for the students to use the learning module, (3) a post-test, and (4) a survey asking for their opinions about the module. The workshop focused on the first two LOs (*Draw arrows* and *Draw products*).

### A2.1.2 Findings

The cohort I study showed learning gains between the pre-test (M = 61%) and the post-test (M = 70%), $t(8) = -2.29$, $p = 0.051$. These results were investigated in greater depth in the Cohort II study (*vide infra*).

Participants used very few strategies, such as mapping or expanding, although the strategies are taught in the module and have been correlated with successful problem solving in organic synthesis questions. Every participant answered Question 1 correctly on both the pre-test and the post-test. Because of this ceiling effect, the researchers replaced that question with a *Draw the arrows* question in the cohort II study.

All the participants completed a survey about their experiences. They replied positively when asked if the module's effect. For example, when answering the question: "Do you think

your use of the organic reaction mechanisms module will have an impact on your success in your course? If so, how?", Participant 1 wrote that "*It is very good for learning the basics*". Participants 3, 5, 6 and 7 all mentioned that the practice questions and instant feedback were useful and the main reason they liked the module. Participant 9 stated that the module was good "*to teach me everything I was behind on*" and Participant 2 mentioned that the module was useful to "*Make sure I understand the basics"*. Table A2.1 shows the common themes recurring from the students' answers in the survey.

*Table A2.1. Students' comments regarding the features of the learning module.*

| Best Features | Worst Features |
|---|---|
| Quick response to answers | Occasional errors in the answers [a] |
| Module interface was easy | Video question sometimes hid key information |
| Good explanations and descriptive videos | Draw the arrow tool was hard to use |
| Self-paced | Mapping questions were hard to see [a] |

[a] The developers have resolved these issues.

A few technical difficulties were reported, such as occasional errors in the feedback, problems logging in, and difficulty with the drawing ability of the learning module. These issues have been addressed by the developers and were not significant barriers to the students' experiences or learning.

## A2.2 Coding scheme

For LO1: *Draw the arrows* questions, each arrow was awarded one point (Figure A2.1). Arrows must start from an electron pair or bond and point to the correct atom or bond. Therefore, Question 1 was worth 4 points, question 2 was worth 6 points, question 3 was worth two points, and question 4 was worth 3 points.

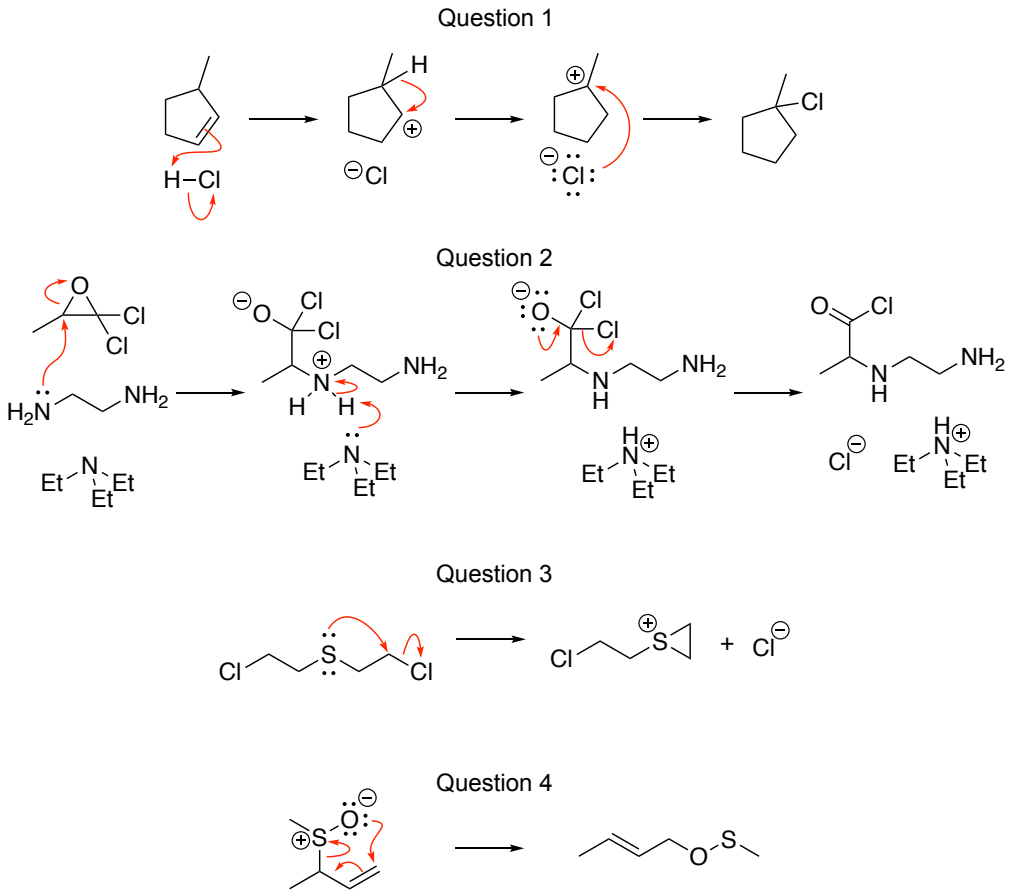
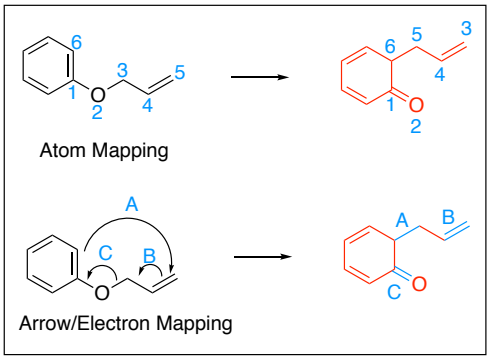



*Figure A2.1. Coding scheme for Draw the arrow questions. Each arrow was awarded 1 point.*

For LO2: *Draw the products* questions, one point was awarded for interpreting each aspect of the electron pushing arrow: breaking and making a bond. For the example question 5, shown in Figure A2.2, arrow A was worth two points: one point was awarded for breaking the □ bond between carbons 1 and 6 and one point for making the bond between carbons 5 and 6. Each response was coded accordingly. All of the other *Draw the arrow* questions were coded in this way (Figure A2.4–20).

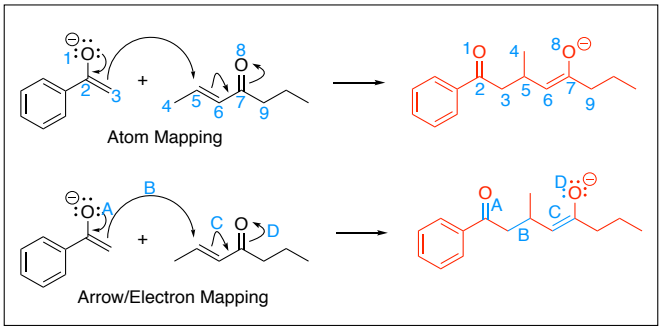
### A2.2.1 Strategies used

Both LO1 and LO2 were coded with the same strategies. Table A2. shows the description associated with each strategy for both types of questions.



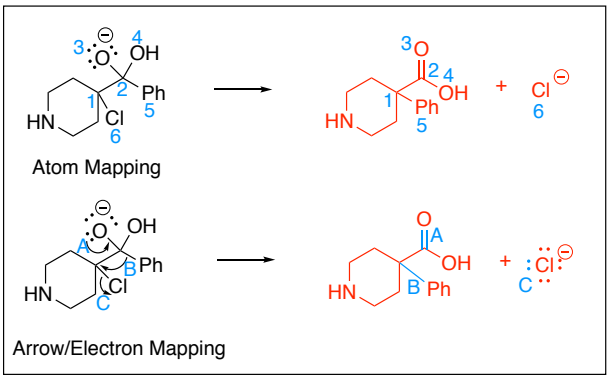

| Action | Bond involved | Arrow | Type of bonds |
|---|---|---|---|
| Break | C1-C6 | A | π |
| Make | C6-C5 | A | σ |
| Break | C4-C5 | B | π |
| Make | C3-C4 | B | π |
| Break | O2-C3 | C | σ |
| Make | C1-O2 | C | π |

*Figure A2.2. Coding scheme for question 5: Draw the products.*



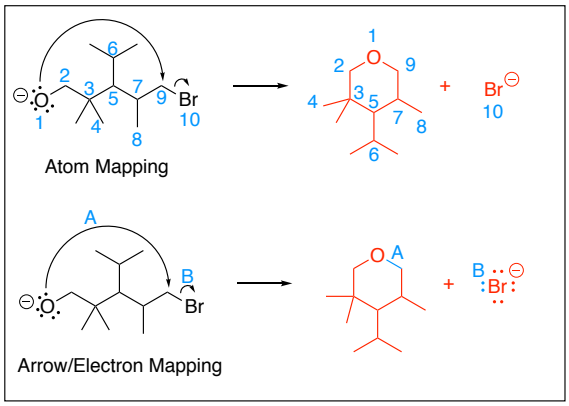

| Action | Bond involved | Arrow | Type of bonds |
|---|---|---|---|
| Make | O1-C2 | A | π |
| Break | C2-C3 | B | π |
| Make | C3-C5 | B | σ |
| Break | C5-C6 | C | π |
| Make | C6-C7 | C | π |
| Break | C7-C8 | D | π |

*Figure A2.3. Coding scheme for question 7: Draw the products.*




| Action | Bond involved | Arrow | Type of bonds |
|---|---|---|---|
| Make | O3-C2 | A | π |
| Make | C1-C5 | B | σ |
| Break | C2-C5 | B | σ |
| Break | C1-Cl6 | C | σ |

*Figure A2.4. Coding scheme for question 6: Draw the products.*




| Action | Bond involved | Arrows | Type of bonds |
|---|---|---|---|
| Make | O1 – C9 | A | σ |
| Break | C9-Br10 | B | σ |

*Figure A2.5. Coding scheme for question 8: Draw the products*

*Table A2.2. Coding scheme for the strategies used.*

| Code | LO1: Draw the arrows | LO2: Draw the products |
|---|---|---|
| Attempted | Must have minimum 1 arrow | Must have any structures |
| Drew non-bonding electrons | Must have at least 1 non-bonding electron pair | Must have at least 1 non-bonding electron pair on the product |
| Expanded the structure | Explicitly drew proton(s) and/or wrote out a C for carbon | Explicitly drew proton(s) and/or wrote out a C for carbon |
| Re-drew the structure | Re-drew all or part of the structure | Re-drew all or part of the structure |
| Mapping | Any attempt to identify atoms in both the SM and product | Any attempt to identify atoms in both the SM and product |

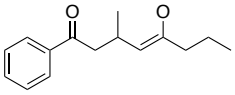
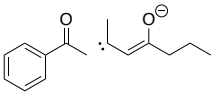
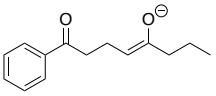
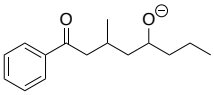
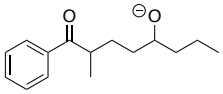
For LO1, each arrow is assigned one of the codes outlined in Table A2.1. The codes were adapted from Flynn and Featherstone's work (2017).

*Table A2.1. Coding scheme for errors in Draw the arrows questions.*

| Code | Definition | Example |
|---|---|---|
| Correct arrow | The arrow is correct |  |
| Missing arrow | The arrow is missing |  |
| Extra arrow | There is an extra arrow to show the step |  |
| Reversed arrow | The arrow starts at the electron deficient site and points to the electron rich site |  |
| Arrow from the atom | The arrow starts at an atom (the arrow would be "correct" if it had started from an electron pair) |  |
| Arrow from the charge | The arrow starts at a charge (the arrow would be "correct" if it had started from an electron pair) |  |
| Wrong arrow | The arrow starts and/or end at the wrong sites |  |
| Vague arrow | The arrow is too vague to determine where it starts or points to |  |
| Did not attempt | The question was left blank | |

For LO2: *Draw the products* several errors were coded for (Table A2.2). The errors were once again adapted from previous work (Flynn and Featherstone, 2017).

*Table A2.2. Coding scheme for the errors of LO2 Draw the products.*

| Error | Definition | Example |
|---|---|---|
| Formal Charge (FC) error | When the structure has either the wrong FC, a missing FC, or an extra FC. The formal charge error must be in relation to the structure they drew. |  |
| Transplanting electrons | Taking electrons and moving them without forming a bond |  |
| Missing/extra atom | There is a missing or extra atom in the molecule drawn |  |
| Missing/Extra bond | There is a missing or extra bond in the molecule drawn |  |
| Placement error | The atoms are not correctly connected (typically ??is misplaced) |  |
| Did not attempt | No answer was provided | |

# Appendix 3: Synthesis of delocalization LO and Exam Analysis (Chapter 5) supplemental information

## A3.1 Which term is appropriate: resonance or delocalization?

Resonance is often introduced as a tool to represent a molecule that cannot be represented by a single Lewis structure (Ogilvie *et al.*, 2018). Several Lewis structures with differences in  and non-bonding electrons are used to portray a single molecule, but the discrete resonance structures do not exist. A molecule is actually a combination all of the resonance structures, which can be represented using the resonance hybrid; however, the hybrid structure does not show the number of electrons that are delocalized.

The International Union of Pure and Applied Chemistry (IUPAC) defines resonance as: "the representation of the electronic structure of a molecular entity in terms of contributing structures…" (McNaught and Wilkinson, 2006).

Kerber sparked a debate when he stated that resonance was a misleading term and should be replaced by the word "delocalization" (Kerber, 2006). He advocated using the word "delocalization" and "delocalization forms" instead of "resonance". Other chemists have opposed Kerber's views and stated that "resonance" is both conceptually and historically more accurate (Jensen, 2006). This type of argument can also be demonstrated in the fact that several advanced organic chemistry textbooks (Fleming, 2009; Clayden *et al.*, 2012) have decided to not use the word "resonance" but instead use the word "delocalization". Those authors also note that delocalization is often referred to as resonance.

According to the IUPAC definition, "resonance" refers to any Lewis structure that contributes to the wavefunctions, however unlikely that structure is (McNaught and Wilkinson, 2006). This would mean that delocalization across sigma bonds are resonance structures and that hyperconjugation would also fall under the definition of resonance (Mullins, 2012). Therefore, resonance structures do not solely differ on placements of π orbitals electrons and non-bonding electrons but include any structure that contributes to the wavefunction. Electrons being shared over multiple atoms in π orbital is actually delocalization (McNaught and A Wilkinson, 2014),

while the Lewis structures associated with the π-system are mesomerism (McNaught and A. Wilkinson, 2014).

## A3.2 LOs and the percentage of questions related to each ILO

Thirty-three LOs were derived from the first coding, which were then validated by experts (Table A3.1) The table also shows the percentage of end-of-chapter questions that were related to each of the individual ILOs. The three of the ten essential LOs did not appear in their entirety in the table; due to be an aggregation of several of the 33 ILOs. Some of the ILO are too fine-grained and has such were added together to form one essential LO, such as LO10 (Reaction) which includes "Draw the nucleophilic aromatic substitution *mechanism including all structures representing delocalization"* and "*Use delocalization to explain regioselectivity"* among others. Therefore, LO10 includes all ILOs that require the concept of delocalization in the context of a reaction. Similarly, LO7 (*Stability*) is all encompassing of all ions, therefore the ILOs related to anion stability and cation stability were binned together. On the other hand, LO2 (*Draw*) requires two of the 33 ILOs to be present to be addressed: "*Draw the resonance structures"* and "*Use curved arrows to show electron delocalization"*. For most of the coding the electron-pushing formalism was implicit in the questions, therefore questions that require to draw resonance structures (without explicitly stating to draw the EPF), were coded for the essential LOs.

*Table A3.1. Percentage of all delocalization questions (N = 1548) related to the 33 ILOs.*

| Intended LOs | Questions, % |
|---|---|
| 29. Predict the product of an electrophilic aromatic substitution (EAS) reaction | 19.9 |
| 1. Draw the resonance structures (ILO2) | 15.4 |
| 27. Describe the effect of resonance on a molecule acidity/basicity (ILO8) | 10.1 |
| 11. Predict whether a molecule is aromatic, anti-aromatic or non-aromatic | 6.3 |
| 4. Identify structures where electron delocalization can occur (ILO1) | 5.8 |
| 24. Classify aromatic substituents as activating or deactivating | 4.1 |
| 31. Predict the product of a nucleophilic aromatic substitution reaction ($S_NAr$) | 3.4 |
| 5. Assess contribution of each resonance structures to the hybrid, including justification (ILO3) | 3.1 |
| 32. Determine if the nucleophilic reaction of a conjugated system gives the kinetic or thermodynamic product | 3.0 |
| 12. Justify why a molecule is aromatic (or anti-aromatic or non-aromatic) using concept of conjugation, planarity, cyclic structures and Hückel's rule (ILO6) | 2.8 |
| 28. Draw the electrophilic aromatic substitution (EAS) mechanism with all structures representing delocalization (ILO10) | 2.7 |
| 14. Use Hückel's rule to determine if a molecule is aromatic, anti-aromatic or non-aromatic | 2.5 |
| 18. Correlate delocalized structures to spectroscopic data | 2.1 |
| 33. Draw the mechanism of a 1,2 or 1,4 conjugate addition | 2.1 |
| 25. Use delocalization to explain regioselectivity (ILO10) | 1.7 |
| 26. Predict electrophilic and nucleophilic sites on molecules (ILO9) | 1.7 |
| 16. Use delocalization to explain bond length | 1.3 |
| 20. Use delocalization to explain cation energy/stability (ILO7) | 1.3 |
| 2. Draw the electron pushing formalism arrows to represent electron delocalization (ILO2) | 1.2 |
| 15. Correlate structure with macroscopic properties, such as light absorption. | 1.2 |
| 17. Explain how delocalization affects bond rotation | 1.2 |
| 8. Identify hybridization of atoms involved in delocalization (ILO5) | 1.1 |
| 13. Use Frost circles to justify aromaticity or anti-aromaticity | 1.0 |
| 23. Use delocalization to explain radical energy/stability | 1.0 |
| 3. Draw the resonance hybrid including the partial charges (ILO4) | 0.7 |
| 19. Explain the effect of delocalization on heat of hydrogenation | 0.7 |
| 30. Draw the $S_NAr$ (nucleophilic aromatic substitution) mechanism with all structures representing delocalization (ILO10) | 0.7 |
| 6. Draw the orbitals involved in electron delocalization | 0.6 |
| 22. Use delocalization to explain arenium energy/stability | 0.6 |
| 7. Draw the MO diagram of a molecules with resonance | 0.3 |
| 21. Use delocalization to explain anion energy/stability (ILO7) | 0.3 |
| 9. Explain the effect of delocalization on the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) | 0.0 |
| 10. Explain $\pi$ orbitals alignment and the molecule planarity required for delocalization to occur | 0.0 |

## A3.3 Example of Taught-Practiced-Assessed (TPA) coding

To code for how a paragraph or section is teaching resonance, each paragraph in sections related to delocalization was assigned an ILO (or multiple) that was being demonstrated in the paragraph. Figure A3.1 shows an example of a section in a textbook. The callout on the right shows the ILOs that were coded for each paragraph. The callout within the text shows the types of recurring themes that were highlighted throughout the coding.

The following series of practice questions would be coded using the 33 ILOs that have been proposed, not just the ten essential ILOs (Table 2). Some ILOs are present in the questions, but not completely, in those cases the ILOs were coded as present but with a note stating why it was not fully addressed.

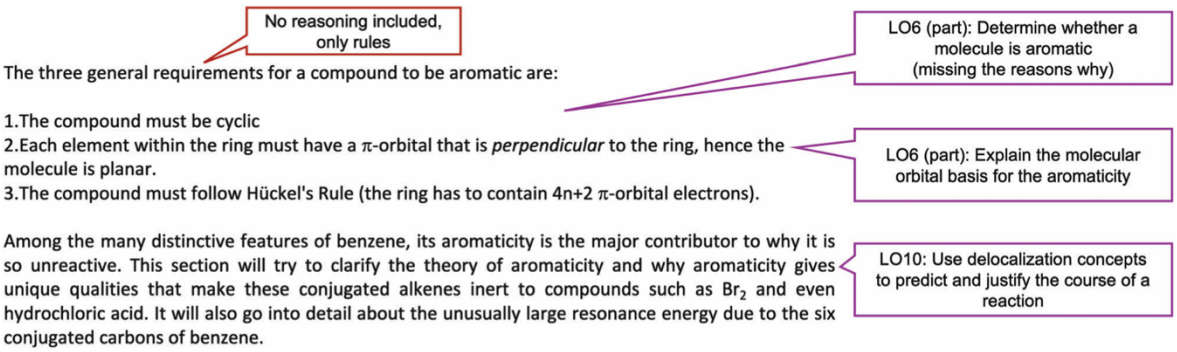
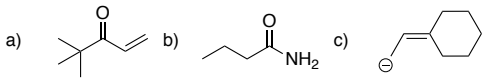
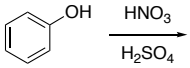
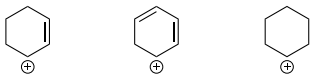



*Figure A3.1. Example of how textbook paragraphs were coded. Excerpt taken from Chemistry Libretexts (2019).*

*Table A3.2. Coding examples of how questions were coded. LOs are listed in Table 1.*

| Question | Code | Explicit or implicit |
|---|---|---|
| Draw the resonance structures for the following molecules circle the major  contributor to the resonance hybrid: | ILO1<br>ILO5 (no reasoning) | ILO1: Explicit<br>ILO2: Explicit |
| Draw the mechanism of the following electrophilic aromatic substitution  reaction | ILO28 (not fully addressed) | Implicit |
| Rank the following molecules in order of increasing stability  | ILO20 (no reasoning)<br>ILO4 | Implicit for both LOs |

## A3.4 Exam analysis coding scheme for RQ1: LO Achievements

Table A3.3 outlined the criteria to achieve a specific LO.

*Table A3.3. General criteria to achieve the LOs*

| LO | General criteria to achieve the LO |
|---|---|
| LO1 | Stated that delocalization can occur or drew resonance structures when not explicitly asked |
| LO2 | Correctly drew the resonance structures, with the proper bonds and charges |
| | Correctly drew the curved arrows to show electron delocalization, i.e., start at the source of electrons and finish are the correct position |
| | There is no change in the sigma bonds between resonance structures |
| | There are no extra or missing structures |
| LO3 | Correctly identified the major contributor or correctly identified the order of contribution to the resonance hybrid |
| | Correctly used the evidence of charge and octet to justify their answers |
| LO4 | Correctly drew the dashed bonds representing electron delocalization |
| | Correctly drew the partial charges (not full charge) |
| LO5 | Correctly identify the hybridization of atoms in delocalized system |
| LO6 | Correctly label the cycles as aromatic, anti-aromatic and non-aromatic |
| LO7 | Correctly identify the most stable structure |
| | Justify their answers by stating the cation is resonance stabilized OR by drawing the resonance structures |
| LO8 | Correctly identify the most acidic proton or basic atom |
| | Justify the answer by stating that a conjugate base can be stabilized by resonance |
| LO9 | Correctly identify the most nucleophilic atom (or electrophilic) |
| | Justify their answers by stating delocalization will lower nucleophilicity OR by drawing the resonance structures |
| LO10 | Correctly identifying that the regio-selectivity of a reaction |
| | Justify their answers by stating delocalization will stabilize an intermediate |

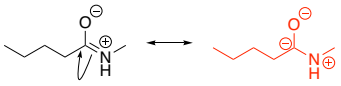
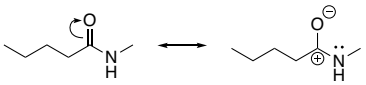
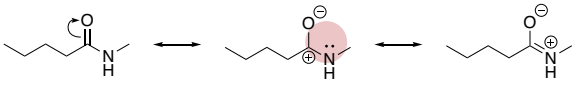
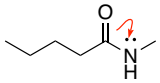
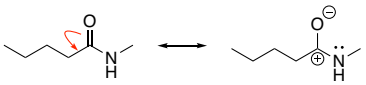
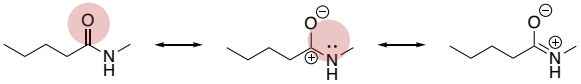
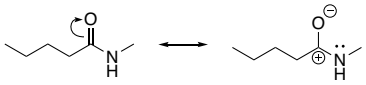
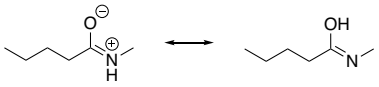
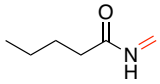
### A3.4.1 LO1 (Identify)

LO1(Identify) was coded as present if an answer contained resonance structures, or it was indicated that resonance structures would be present.

### A3.4.2 LO2 (Draw)

LO2 (Draw) was assessed in five questions and the coding scheme can be seen in Table A3.4.

*Table A3.4. Coding scheme for LO2*

| | Description | Example |
|---|---|---|
| Extra structure | There is an extra structure (either extremely minor, or wrong). If the same structure is there multiple time, it does not count as extra |  |
| EPF | **Correct** – The curved arrows start from electrons and point to the appropriate location |  |
| | **Missing** – one or more of the curved arrows is missing (however at least 1 is drawn) |  |
| | **Reversed** – curved arrow starts at where it should end and vice-versa |  |
| | **Extra** – there is an extra arrow present | Not seen |
| | **Wrong** – the curved arrow starts and ends in wrong location |  |
| | **None** – No curved arrows were drawn at all |  |
| | Double headed arrow |  |
| | Reaction arrow |  |
| | Equilibrium arrow |  |
| Formalism | No arrows between the structures |  |
| Overall errors | **Reaction** - Answer showed a reaction occurring, |  |
| | **Impossible structure** – Answer had structures that are impossible i.e., 5 subs to a carbon, |  |

### A3.4.3 LO3 (Assess)

For LO3 (Assess) the students order of contribution to the hybrid was coded, as well as which evidence they used (*i.e.*, octet or charge for all three structures). We then used the students' mode of reasoning as outlined in Sevian and Talanquer (2014)[5] to code their reasoning (Table A3.5).

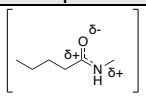
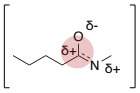
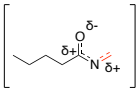
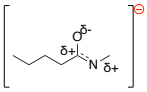
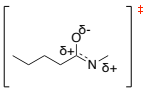
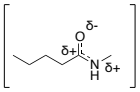
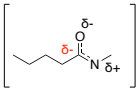
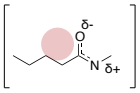
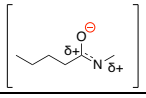
*Table A3.5. Coding scheme of student reasoning for LO3 (Assess)*

| Mode of reasoning | Code | Explanation | Example |
|---|---|---|---|
| Descriptive | D | Describes the octet/charge | See expected answer |
| Relational | R | Links octet/charge to contribution to the contribution | "It is the major contributor because all atoms have a full octet and no charge." |
| Linear | L | Uses octet OR charge to explain the contribution and explains WHY the concept used is relevant | "It is the major contributor because all atoms have a full octet, which |
| Multi-component | M | Uses octet and charge to explain the contribution and explains WHY the concepts used are relevant | "It is the major contributor because all atoms have a full octet, which would mean all atoms have filled orbitals, and not charge, meaning there is no build-up of electron density on one atom" |

### A3.4.4 LO4 (Hybrid)

We coded the following errors for the resonance hybrid (Table A3.6). All examples of the errors are provided for the correct resonance structures for Question 1.

*Table A3.6. Coding scheme for LO4*

| | Description | Example |
|---|---|---|
| Dashed bonds | **Correct:** The dashed bonds were correct for the resonance structure draw |  |
| | **Missing** – One of the dashed bond was missing |  |
| | **Extra** – there is an extra dashed bond |  |
| Formalism | Positive or negative charge |  |
| | Transition state |  |
| Charge | **Correct** – The correct charges were present |  |
| | **Wrong Charge** – Answer had a positive charge instead of a negative and vice versa |  |
| | **Missing/Extra charge** - the answer either lacked a charge or had too many from the structures they had drawn |  |
| | **Full charges** – The hybrid had full charged instead of partial |  |

### A3.4.5 LO5 (Hybridization)

We coded the answers for the hybridization they provided. Since no reasoning was required, we found that it was all we could capture.

### A3.4.6 LO6 (Aromaticity)

We coded the answers for the type of cycle they provided. Since no reasoning was required, we found that it was all we could capture. The strategies used was also coded (Strategies section below)

### A3.4.7 LO7 (Stability)

We coded the answers for which molecule they selected. Since no reasoning was required, we found that it was all we could capture. The strategies used was also coded (Strategies section below)

### A3.4.8 LO8 (Acid-base)

For LO8 (Acid-base) the claim (most acidic proton) was coded. The written explanation was then coded for the concepts used (Table A3.7). Following this, each concept that is linked, via linking words (but, therefore, as such, for example, because) or symbols (=, <, >) was be coded.

*Table A3.7. List of concepts coded for LO8 (Acid/base) - the acid–base question*

| Code | Concept | Explanation |
|---|---|---|
| AS | Acid strength | Mentions weak or strong acid |
| BS | Base strength | Mentions weak or strong base |
| B | Conj. base stability | Mentions the relative stability of the conj. base |
| R | Resonance | Mention the concept of resonance |
| I | Induction | Mentions the concept of induction |
| S | Steric hinderance | Mention of hinderance, or bulky |
| P | Proximity | Mention that it is closer to O |
| EN | Electronegativity | Mentions that O is more EN or that it can accommodate electrons better |
| E | Electrons | Mentions electrons |
| C | Charge | Mentions the charges that can be present. |
| O | Other | |

Mode of reasoning for the written part of the question will be viewed holistically using the following rubric, which was adapted from [5]

*Table A3.8. Mode of reasoning for LO8 (Acid–base) – question 2*

| Mode of reasoning | Explanation | Concept link | Example |
|---|---|---|---|
| Descriptive | Describes some aspect of the evidence | None | "It is closer to the oxygen"<br>"There is resonance in A" |
| Relational | Links two concepts together to explain the phenomenon | Acid strength → Stability<br>Base strength → Stability<br>Base strength → Acid strength | "Base A is more stable and therefore it's a stronger acid" |
| Linear | Creates a linear-causal relationship i.e., explains WHY resonance affects the stability/acidity | resonance → stability → base strength | "This is a weaker base because the base is more stable due to resonance." |
| Multi-component | Uses multiple linear causal relationship | resonance → stability ← induction,<br>↓<br>base strength | "Resonance stabilizes the conjugate base and therefore it is a weaker base. The conjugate base is also stabilized via induction with the oxygen. Since the conjugate base is stabilized by resonance AND induction it is a weaker base/acid." |

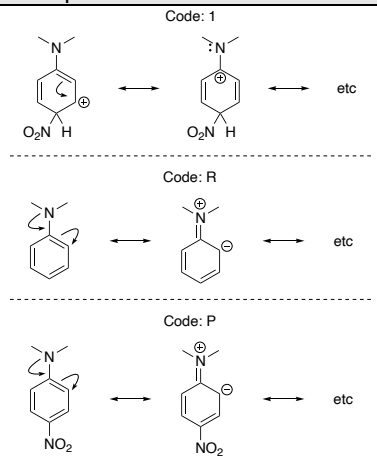
### A3.4.9 LO9 (E+/Nu)

We coded the answers for which molecule/atom they selected. Since no reasoning was required, we found that it was all we could capture. The strategies used were also coded (Strategies section below).

### A3.4.10 LO10 (Reaction)

For LO10 (reaction) coded for which Regio isomer students selected as the major product, as well as which structures were drawn.

*Table A3.9. Coding scheme for LO10 (Reaction) question - EAS questions*

| | Description | Example |
|---|---|---|
| Drew delocalization on the right structure? | 1 = Drew resonance structures for the arenium<br>P = Drew resonance structure for the product<br>R = Drew resonance structures for the reactant |  |
| Which resonance structure was drawn? | P = para<br>O = ortho<br>M = meta<br>Mp = meta and para<br>Op = ortho and para<br>Om = meta and ortho<br>Mop = all three<br>Empty = did not draw resonance OR something other | Coded for which structures were present on the exam |
| Structure missing | X – The letter associated with the missing structure<br>*At least one other structure was present | |

While not necessary some students included extra information. The following information was coded as present or not (Table A3.10. Extra information coded for EAS type questions).

*Table A3.10. Extra information coded for EAS type questions*

| Concept | Description |
|---|---|
| Steric hinderance | 1 = Mentioned steric,<br>Empty = no mentions |
| Delocalization/ resonance | 1 = Mentioned resonance,<br>Empty = no mentions |
| *Stability* | *1 = Mentioned stability,*<br>*Empty = no mentions* |
| *O/P director* | *1 = Mentioned o/p director,*<br>*Empty = no mentions* |

### A3.4.11 Strategies

While not required many students used strategies in their answer. The strategies were coded for their presence.

*Table A3.11. Coding for the strategies used*

| Strategies | Description |
|---|---|
| Visualizing electrons | 1 = Drew/mentioned resonance or drew arrows exclude reaction mechanism |
| | Empty = strategy not seen |
| Listing properties | 1 = Listed p$K_a$ or hybridization |
| | Empty = strategy not seen |
| Expanding the structure | 1 = Drew LP or implicit atoms, |
| | Empty = strategy not seen |
| Listing rules | Listed rules in the margin, |
| | Empty = strategy not seen |
| Others | 1 = any other strategy |
| | Empty = no other strategy was used |

## A3.5 Expected answer for each question analyzed

Figure A3.1 shows the expected answer for Question 1. The answer would get a perfect score and is aligned with how this question type was taught in the course.

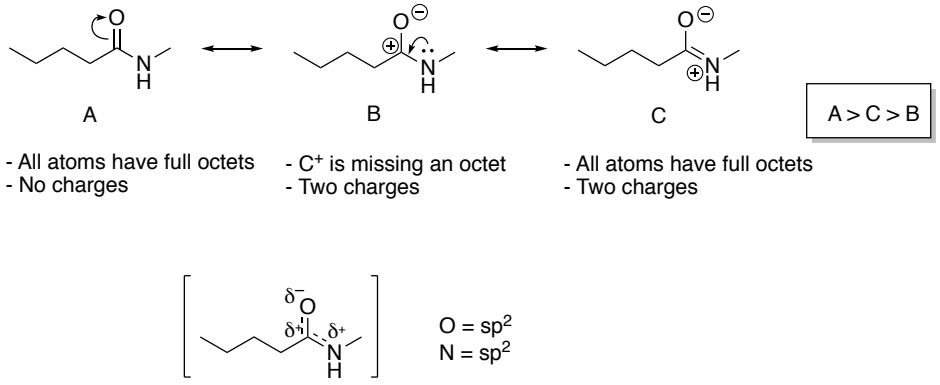



*Figure A3.1. Expected answer for the Question 1.*

Figure A3.2 shows the expected answer for Question 2. The answer would get a perfect score and is aligned with how it was taught in the course.

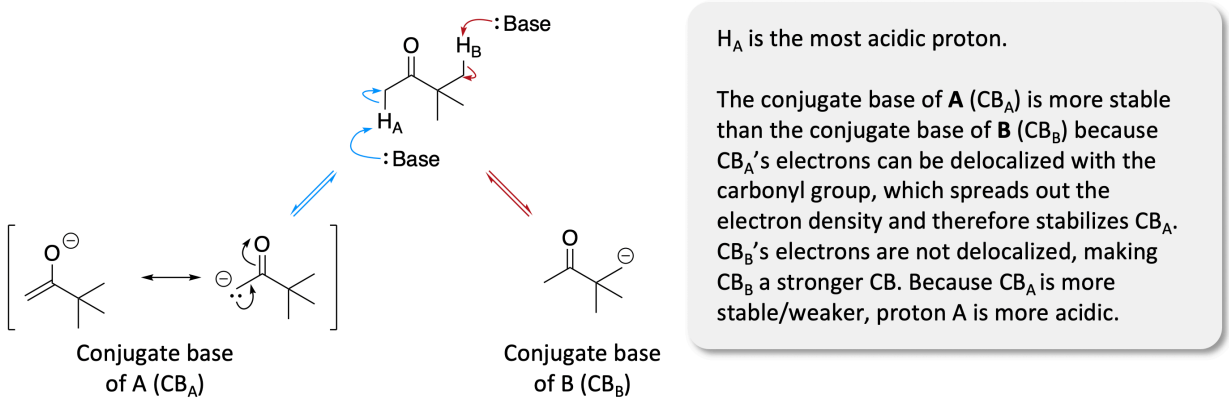



*Figure A3.2. Expected answer for Question 2.*

Figure A3.3 shows the expected answer for Question 3. The answer would get a perfect score and is aligned with how it was taught in the course.

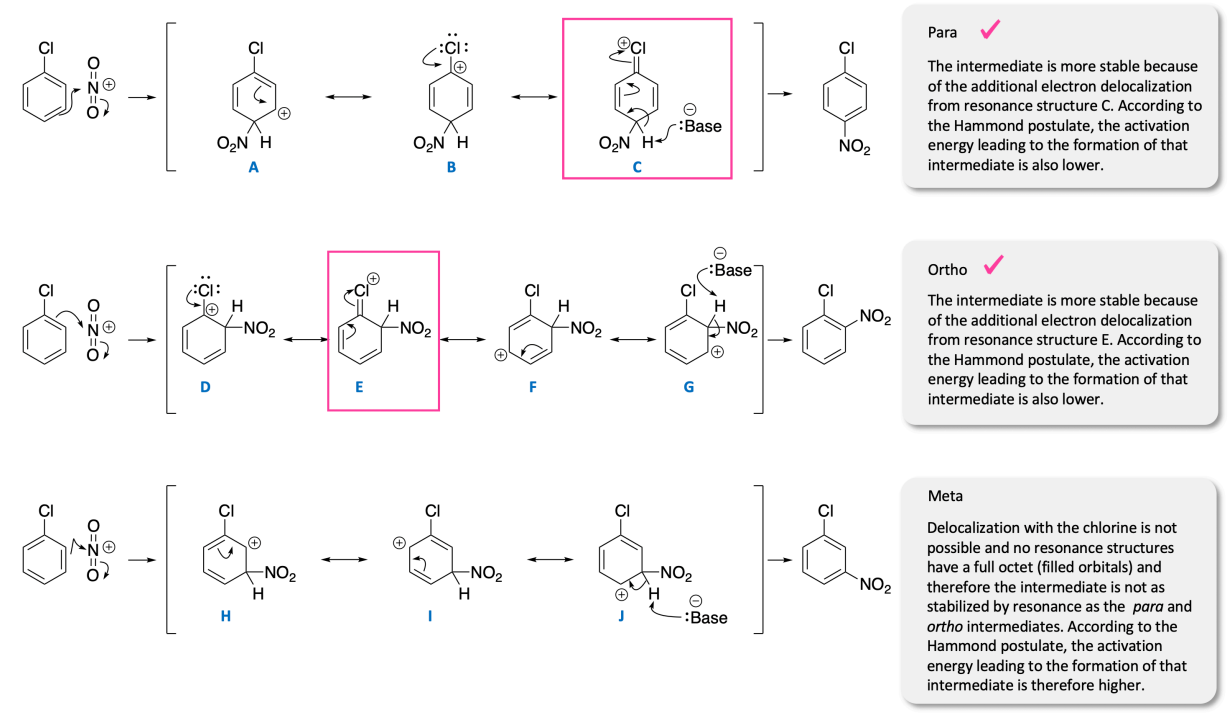



*Figure A3.3. Expected answer for Question 3.*

Figure A3.4 shows the expected answer for Question 5. The answer would get a perfect score and is aligned with how it was taught in the course.

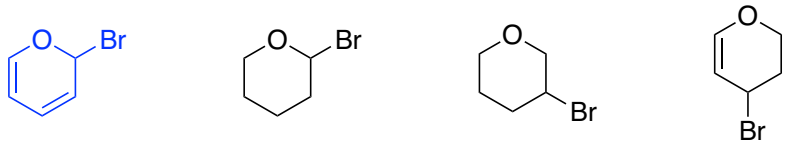



*Figure A3.4. Expected answer for Question 4. The correct answer is in blue.*

Figure A3.5 shows the expected answer for Question 5. The answer would get a perfect score and is aligned with how it was taught in the course.

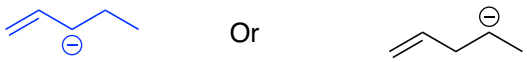



*Figure A3.5. Expected answer for Question 4. The most stable ion is highlighted in blue.*

Figure A3.6 shows the expected answer for Question 6. The answer would get a perfect score and is aligned with how it was taught in the course.

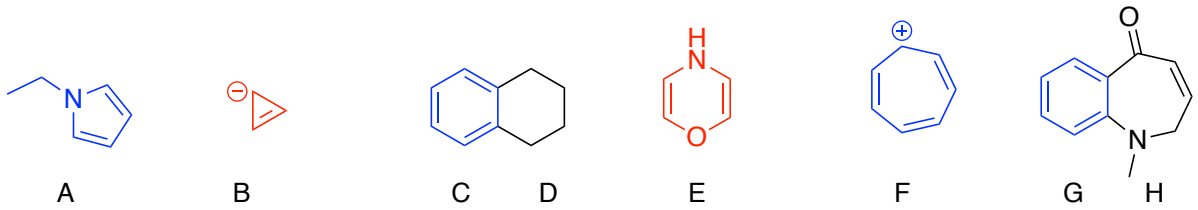



*Figure A3.6. Expected answer for Question 5. Blue = aromatic, Red = anti-aromatic, black = non-aromatic.*

Figure A3.7 shows the expected answer for Question 7. The answer would get a perfect score and is aligned with how it was taught in the course.

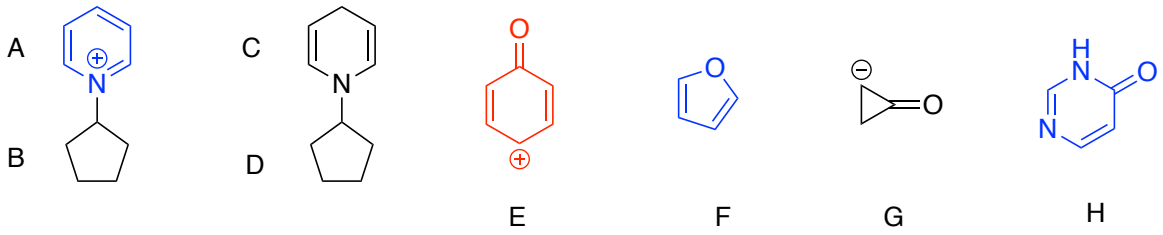



*Figure A3.7. Expected answer for Question 7. Blue = aromatic, Red = anti-aromatic, black = non-aromatic.*

Figure A3.8 shows the expected answer for Question 8. The answer would get a perfect score and is aligned with how it was taught in the course.

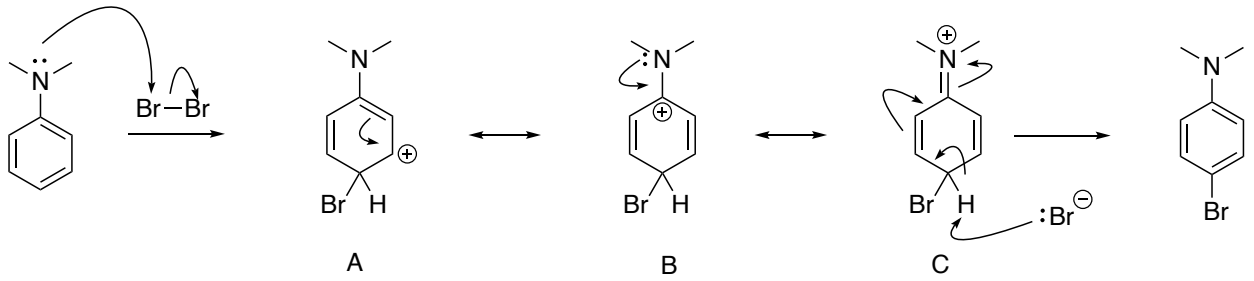



*Figure A3.8. Expected answer for Question 8.*

Figure A3.9 shows the expected answer for Question 9. The answer would get a perfect score and is aligned with how it was taught in the course.

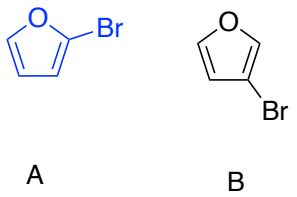



*Figure A3.9. Expected answer for Question 9. The two structures are the expected structures to be drawn. A (blue) is the major product formed.*

Figure A3.10 shows the expected answer for Question 10. The answer would get a perfect score and is aligned with how it was taught in the course.

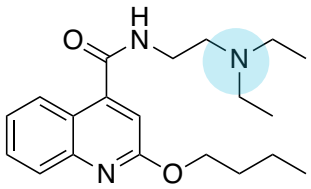



*Figure A3.10. Expected answer for question 10. The most basic atom is in blue.*

Figure A3.11 shows the expected answer for Question 11. The answer would get a perfect score and is aligned with how it was taught in the course.

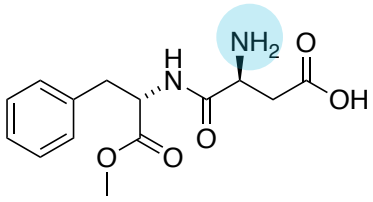



*Figure A3.11. Expected answer for question 11. The most nucleophilic atom is in blue.*

Figure A3.12 shows the expected answer for Question 12. The answer would get a perfect score and is aligned with how it was taught in the course.

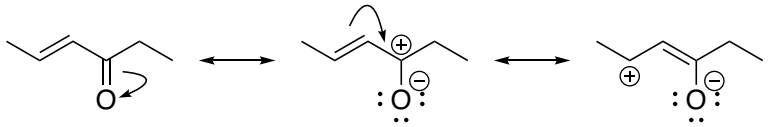


*Figure A3.12. Expected answer for question 12.*

## A3.6 Reliability for coding of the exams

Table A3.12 shows the inter-reliability values for each question. A percent agreement of over 80% and a Krippendorf $\alpha$ value above 0.7 are considered acceptable interrater reliability values.[6] The sections highlighted are the ones that did not meet the threshold on the first round and so the raters met and discussed. After changes to the coding scheme were made, a different subset of exams was coded and compared. The values shown are after the first round of coding for those are in paratheses.

*Table A3.12. Reliability of coding. Number in parentheses are the unsuccessful first rounds.*

| | | LO Achievement | | Strategies and Errors | | Reasoning | |
|---|---|---|---|---|---|---|---|
| Q | LO | % Agreement | Krippendorf α | % Agreement | Krippendorf α | % Agreement | Krippendorf α |
| | LO2 | 100% | -- | 97% | 0.74 | -- | -- |
| | LO3 | 93% | 0.69 | -- | -- | 86% | 0.91 |
| | LO4 | 100% | -- | 98% | 0.69 | -- | -- |
| Q1 | LO5 | 100% | -- | -- | -- | -- | -- |
| | LO1 | 100% | -- | -- | -- | -- | -- |
| | LO2 | 89% | 0.92 | 96% | 0.94 | -- | -- |
| Q2 | LO8 | 100% | -- | -- | -- | 81% (69%) | 0.75 (0.54) |
| | LO2 | 100% | -- | 93% | 0.89 | -- | -- |
| | LO3 | 100% | -- | -- | -- | -- | -- |
| Q3 | LO10 | 100% (60%) | (0.70) | -- | -- | -- | -- |
| Q4 | LO6 | 100% | -- | 97% | 0.81 | -- | -- |
| Q5 | LO7 | 99% | 0.99 | 91% | 0.83 | -- | -- |
| Q6 | LO6 | 100% | -- | 93% (80%) | 0.71 (0.65) | -- | -- |
| Q7 | LO7 | 93% | 0.93 | 94% | 0.82 | -- | -- |
| | LO2 | 100% | -- | -- | -- | -- | -- |
| Q8 | LO10 | 94% (71%) | 0.77 (0.52) | 100% | -- | -- | -- |
| Q9 | LO10 | 100% | -- | 94% | 0.76 | -- | -- |
| Q10 | LO8 | 95% | 0.81 | 96% | 0.75 | -- | -- |
| Q11 | LO9 | 100% | -- | 98% | 0.79 | -- | -- |
| Q12 | LO2 | 100% | -- | 95% | 0.83 | -- | -- |

## A3.7 Assessing LO2 when the starting material was incorrect

For Question 2, only the answers of students that had identified that delocalization was present (n = 155) were analyzed for LO2. This question required students to show a deprotonation first and 27% of the students drew the product of the deprotonation with positive charges (Figure 13), despite the curved arrows of the acid–base mechanism showing electrons moving on the carbon atom. For this research, the resonance structures were coded as correct if the students drew the appropriate structure for the product they obtained.

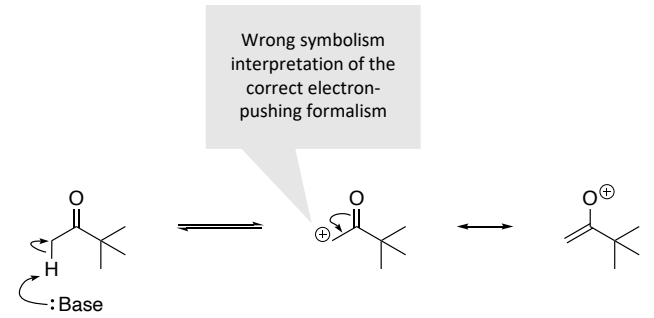



*Figure A3.13. Question 2: Common error (18%) of drawing the wrong charge on the resonance structures, despite drawing the correct curved arrows.*

Similarly, while Question 3 explicitly asked for resonance structures within the mechanism; however, many students (41%) did not draw any resonance structures. For the analysis of LO2, we only included answers that contained resonance structures (N = 105). For the purpose of coding the achievement of this LO (ability to draw resonance structures) we accepted any correctly drawn structures. For example, if a student drew the *meta* intermediate rather than *ortho* and *para* and drew the correct resonance structures (for *meta*), their answer was considered as having achieved LO2 (*Draw*).

## A3.8 Type of answer for reasoning in question 3

The students answered the question by either listing the evidence in a tabular style or by writing out the answer in a paragraph (Figure A3.14). 97% of the students who wrote their answer in a tabulated (or list) format showed descriptive reasoning, as opposed to 80% of the students who used a paragraph format. Relational reasoning was used in 18% of the responses from paragraph answers, as opposed to 1% from table format.

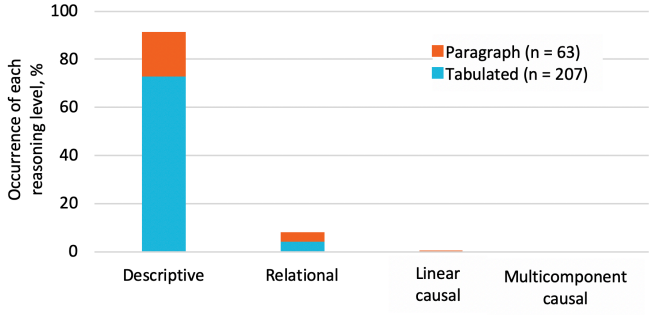



*Figure A3.14. Levels of reasoning in Question 1 (resonance), N = 280.*

We also found a relationship between the mode of reasoning and the format students used to write their answers, with lower reasoning demonstrated most often by students who

summarized their ideas in a table. Causal reasoning could be implied by how the table/list linked the structures to the evidence; however, we decided not to make these assumptions in our coding.