UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Study Behaviors, Problem-Solving, and Exam Design in Organic Chemistry

Permalink
https://escholarship.org/uc/item/25s4t41d

Author
Brando, Beatriz A

Publication Date
2019

Peer reviewed|Thesis/dissertation

eScholarship.org
Powered by the California Digital Library
University of California

Study Behaviors, Problem-Solving, and Exam Design in Organic Chemistry

By

Beatriz A Brando

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Science and Mathematics Education

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Angelica M. Stacy, Chair
Professor Anne M. Baranger
Professor Marcia C. Linn
Professor Sophia Rabe-Hesketh

Spring 2019

©Copyright 2019
Beatriz A. Brando
All rights reserved

Abstract

Study Behaviors, Problem-Solving, and Exam Design in Organic Chemistry

by

Beatriz A Brando

Doctor of Philosophy in Science and Mathematics Education

University of California, Berkeley

Professor Angelica M. Stacy, Chair

Organic Chemistry is considered by many to be the quintessential “weeder course”, widely identified in the premed and STEM persistence literature as a gatekeeper to students interested in STEM or health fields (Barr, Gonzalez, & Wanat, 2008; Lovecchio & Dundes, 2002). There has therefore been considerable interest in determining ways to improve students’ outcomes in organic chemistry courses. While redesigning curriculum and developing interventions is an important step in this process, it is also important to better understand what helps students to do well in organic chemistry.

In this dissertation I utilize a mixed-methods study design to explore how students’ study behaviors and problem-solving strategies are associated with student performance in a second-semester organic chemistry course for non-majors, Organic Chemistry 2. This dissertation begins by exploring how study behaviors are associated with overall performance but then interrogates differences in the relationships between different types of questions (mechanism and predict-the-product) and between questions with different relationships to the curriculum (problem-type and exercise-type questions). It finally conducts a more detailed exploration on the differences between student problem-solving strategies on problem-type and exercise-type questions. In order to explore these issues, the goal of data collection was to provide a comprehensive understanding of the entire course. A survey based on the resources available in Organic Chemistry 2 and students’ study behaviors associated with them was developed and utilized. In-class observations, student exams, course materials, and exam scores were also collected. A subset of students was also interviewed using a think-aloud protocol based on questions from their exams.

Survey results were analyzed using hierarchical linear modeling and multiple linear regression. Results showed that students’ overall exams scores were positively associated with studying more than 10 hours, active note-taking, and testing their understanding when working on practice problems. Exam scores were negatively associated with passive engagement in group-work. However subsequent analyses showed that different categories of questions (mechanism and predict-the-product) are associated with very different strategies from each

other and overall exam scores. Notable differences were also seen between problem-type and exercise-type questions even within the same category of question. All these results indicate that determining what study behaviors are associated with higher scores in Organic Chemistry is more complicated than might be initially expected and that advice to students on how to study should consider the type of questions that students are struggling with. It also indicates that instructors should think critically about more than just the content being tested when designing exam questions.

Interviews were analyzed using thematic coding centered on student decision making and reasoning strategies. Results from initial thematic coding were finalized into activity logs which were then tabulated to better compare the reasoning strategies used in the problem-type and exercise-type questions and their outcomes. Results showed that problem-type and exercise-type questions are associated with different approaches and reasoning strategies. In particular the use of chemical reasoning seems to be important for performance on problem-type questions but other reasoning strategies can be just as effective in exercises. This chapter then discusses how chemical reasoning is undervalued by students and how exercise-type questions may not measure students understanding of chemical concepts.

Table of Contents

Acknowledgements

Graduate School has been a long and winding road and I'm so grateful to the many people who have helped and supported me along the way. It will be almost impossible to thank and acknowledge everyone who helped me make it to the finish line but I'm going to make my best attempt.

I'd like to acknowledge the amazing undergraduate researchers that I've worked with on this project. Janelle Abela, John Della Rosa, Karissa Lapuz, Edar Li, Jennifer Liu, and Kyler McKessy this dissertation would not exist without you all. Thank you for your hard work collecting and analyzing this data. Thank you also for your passion, wonderful ideas, silly jokes, and for your patience with my many rants about education. Jenny Hackleman, thank you for your help as I was first developing this project. Your passion and ideas inspired me both in my education research and in my desire to make more ethical food choices.

To my research advisors, Anne Baranger and Angelica Stacy, thank you for taking a chance on me and supporting me in transferring into the SESAME program. I'm so grateful for this experience and for all I've learned. Thank you for all your support and mentoring over the years and for encouraging me in developing and executing this exciting project.

To my dissertation committee members: Marcia Linn, you have been there for every oral exam I've ever taken in SESAME and you've helped make every one of them an opportunity for growth and learning. Thank you for all your feedback and support throughout my PhD career. Sophia Rabe-Hesketh, I learned all of the quantitative analysis methods that I used in this dissertation in one of your classes. Thank you for everything you've taught me and for all the feedback and support you've given me on my quantitative models.

To Anjum Fitch and Lucia Briggs, thank you all your help collecting data for my research. Anjum thank you also for keeping me fed and being my Pokemon Go buddy.

To Lynn Keithlinn, Kate Capps, and Rosa Garcia, thank you for all your help in figuring out the intense amount of paperwork that graduate school require. Thank you also for looking out for me and for all your help and encouragement.

To Michelle Douskey, my experiences as a GSI with you were what inspired me to pursue a PhD in Education. Thank you for mentoring me and supporting me throughout my time at Berkeley (and for feeding me a lot of chocolate).

To MaryAnn Robak, thank you for welcoming me into your classroom and for countless conversations about organic chemistry education. Your passion for improving chemistry education is inspiring.

To everyone at the CTL, especially Yukiko Watanabe and Richard Freishtat. I learned so much about education and research from you all, but most of all I appreciated the support that you gave me professionally, personally, and in my research. Thank you for listening to many, many hours of rants on my dissertation research and for all your feedback. This dissertation wouldn't have been possible without you.

To Andrew Eppig for help and advice on data analysis. To Suzanne Stone for help and advice on my IRB, and for always being so responsive to questions and amendments.

To Ahamindra Jain, my first college research advisor, who taught me that how we train, teach, and support people in the sciences matters. To Gregg Tucci, who's support of my career continued even after I graduated from Harvard and who remains for me an example of what mentorship and education should look like.

To the graduate students in the chemistry education research group, past and present. Hearing about your work and getting your feedback on mine has helped me to grow as a researcher and a person. Particularly thanks to Laura Armstrong for providing so much support and feedback. Half my gChat is probably messages to you asking your opinion on my research. Also, to Geri, my fellow dissertation writer, and Max my buddy in the organic chemistry problem solving (thanks for paper recommendations and feedback). Thanks also to Sara, Erin, and Michelle for all their advice. Particular thanks to the SESAME Street google hangout for way too many emojis and BuzzFeed quizzes.

Thanks also to my friends in the GSE for all the support and camaraderie. Special shout out to my cohort-mate Beth for coffee dates, research advice, and wine trips. To Renee for being an awesome roommate. To Kyle for encouraging us to build up a community. And to the GSE choir for always being a warm and friendly space.

To all those who work in and around the UC Berkeley campus who have supported me throughout all of my time here: Carl, Latrosse and Roy in the mail room, the wood shop guys, James at the glass shop, Sou with the custodial staff, the employees at various restaurants and coffee shops (particularly the staff at Party Sushi who likely saw me more often than many of my friends), and to many shuttle drivers and BearWalk escorts who made sure I didn't get mugged after a late night of work and were patient, captive audiences as I talked about my research.

To Tarana Wireless, for letting me write a lot of this dissertation in their Berkeley office and for letting me raid their snacks and drinks.

To my wonderful church communities who have supported me through this whole process. To the staff and parishioners at Newman Hall, particularly Father Bill whose kindness and advice kept me going even in hard moments, Colleen who's a force of nature and made choir a wonderful and supportive space for me at Newman, and Father Ivan for his kindness and support. Also, to the wonderful priests, brothers, and parishioners of St. Albert's. I'm so grateful for my community there and all your prayers and encouragement.

To all my friends that started with me in Chemistry and have continued to support me to this day. To Peter for pie and League of Legend tournaments. To Stephanie for puzzling and car trips. To Tish for fantasy novels and handmade jewelry. To Jen for delicious cakes and hilarious stories about accidentally dating magicians. To Ashley for steampunk balls and Holi. To Cory for hiking poles and education conversations. To Tristan for steampunk goggles and bass guitars. To Kate for Facebook stalking, gewürztraminer, and roomie bonding. And to Mike, for breakfasts, dinners, wine, liturgy of the hours, meditation, being a lector at my wedding, and listening to me rant about my research for years. I never would have made it without you.

To my wonderful friends from high school who even though far away have always been my cheerleaders and supporters. To Laura for silly pictures, crazy conversations, and comforting me times when I cried on the phone. To Vivian for Dirk Gently's Holistic Detective Agency, Scorpio Races, and long gchat conversations. I love you ladies and never would have made it through this without you.

To my wonderful family. To my brother Alberto for his support and comfort when I felt overwhelmed. To my great uncle Joaquin for his enthusiastic support throughout grad school. My in-laws Grace and Paul, Erica and Andrew for all their kind support. And to my nieces Nora and Claire whose videos and pictures have cheered me up even when research drove me batty.

In particular though I want to thank my parents though who have been my rock throughout my entire graduate school experience (and throughout my entire life). You were there for me in dark moments and in moments of joy. Thank you Mami for always reminded me to see the beauty in the world and in my work and thank you to Daddy for encouraging me to stay strong and for appreciating the technicalities of the data analysis in my work. Thank you also Mami for helping me prepare for my qualifying exam, and Daddy for teaching me about Gantt charts and online transcription software.

To Mark, you're the love of my life and the best part of my PhD. Thank you for all your support, for listening to many, many rants about education theory, and for accepting the insanity that was writing this dissertation. Thank you also for never letting me give up; I couldn't have done this without you. I'm so lucky that you're my husband.

Finally, to all the students who participated in this research. Thank you for trusting me with your stories and experiences; I love being able to learn from you and hope to live up to your trust in me.

Chapter 1: Introduction and Research Questions

Introduction

Organic chemistry is in many ways the quintessential “weeder” course. It’s a notoriously difficult course required by various STEM majors and its inclusion in both the MCAT and as a medical school prerequisite means that many students consider it imperative to receive high grades in the course. A reputation for difficulty alone would not necessarily be of concern, but there exist several studies which indicate that organic chemistry is not only difficult but also a key factor in students’ decisions to persist in STEM/Health fields.

Improving STEM persistence, particularly for women and underrepresented minorities (URM), remains an important goal in the field of higher education. Studies exploring student STEM persistence have consistently shown that URM students have higher attrition rates in STEM fields than white students (Hurtado et al., 2007). This has also historically been true of female students, though studies show that there has been some improvement in recent years (Miller & Wai, 2015). Several factors have been found to be influential in student persistence. Seymour & Hewitt (1997), for example, in their large ethnographic work identified the following areas: initial motivations for their choice of the field, students’ high school preparation (both in skills, habits, and their relationships with their instructors), the ‘hardness’ of STEM classes, feeling weeded out by the curriculum, the culture of STEM, and how STEM classrooms are taught. Experiences in STEM classes is a dominant theme throughout this work, and throughout the work chemistry (and in particular organic chemistry) was consistently brought up as a class than influenced students to switch to other majors.

This finding is consistent with work that investigated the persistence of premedical students. In their study of premed students at Stanford University (2008) and later at UC Berkeley (2010), Barr et al found that students consistently identified their experiences in chemistry courses as influential in their decisions to leave the premed track (Barr et al., 2008; Barr, Matsui, Wanat, & Gonzalez, 2010). Lovecchio and Dundes, also identified students’ experiences in organic chemistry as influential in the persistence of premed students (Lovecchio & Dundes, 2002). In particular Lovecchio and Dundes found that low performance in organic chemistry often led students to give up on medical aspirations. While more work needs to be done on the reasons why organic chemistry has such an impact on STEM majors and premedical student, what is obvious is that students’ experiences and performance in organic chemistry courses is incredibly influential. Since there exists a relationship between organic chemistry performance and STEM persistence it becomes therefore important to explore the factors that influence student performance in organic chemistry coursework.

Thankfully the field of chemistry education can provide us with some insights into factors that influence student performance in their university chemistry coursework. Research in organic chemistry education has been predominantly dominated by intervention studies. This includes strategies like Peer-Led Team Learning (PLTL) (Tien, Roth, & Kampmeier, 2002) and active learning (Black & Deci, 2000). These approaches tend to emphasize student-centered learning and incorporate some level of group work, and findings indicate a connection between student-centered pedagogy and student performance, motivation, and engagement.

Work has also been done on other factors that influence students' performance in organic chemistry. This work has been approached from a variety of directions, however three areas that have been consistently identified are students' **prior knowledge, study behaviors/motivations,** and **problem-solving approaches.** Study behaviors/motivations and problem-solving approaches were considered of particular interest in this work since they represent areas potentially under students' conscious control but which can also be influenced by instructor pedagogical choices.

Considering the large emphasis usually placed on studying outside of class at the university level it is not surprising that the nature of their studying would influence students' performance in organic chemistry. Work such as that of Lynch and Trujillo (2011) and Zusho et al. (2003), utilized the Self-Regulated learning model (SRL) have consistently identified self-efficacy and task management as being correlated with student performance (Lynch & Trujillo, 2011; Zusho, Pintrich, & Coppola, 2003). Other work such has shown that higher level strategies such as organizing and rearranging instructional materials, and front-loading studying were both correlated with performance (Szu et al., 2011). Finally work using the Approaches to Learning framework found that students who used higher level and more metacognitive strategies, even in their interactions with others, tended to perform better in general chemistry courses (Sinapuelas & Stacy, 2015).

The field of chemistry education has also explored the problem-solving strategies that students engage in and whether they are effective in helping students perform better. While dominated by smaller descriptive qualitative studies, this work has still shown clear overarching themes. One overarching theme has been comparing students who problem-solve algorithmically to those who problem-solve conceptually. Nurrenbern and Pickering, for example, identified that there is a difference between problem solving and conceptual understanding, and that it was problematic to assume that they were interchangeable (Nurrenbern & Pickering, 1987). Researchers have also conducted work on the related area of reasoning strategies (Christian & Talanquer, 2012; Kraft, Strickland, & Bhattacharyya, 2010). This work has compared rules-based reasoning, case-based reasoning, model-based reasoning, and in some work symbol-reasoning and often found that higher-order strategies such as case-based and model-based reasoning are more associated with better performance.

Purpose

The current literature on organic chemistry education has shed light on the impact pedagogical practices, study strategies/motivations, and problem-solving strategies can have on student performance. What has been missing so far, however, is work that explicitly acknowledges and addresses the nature of the assessments that are used to measure student performance. Namely the use of summative exams that are the primary form of assessment in large undergraduate STEM courses such as organic chemistry. Most of the literature on organic chemistry performance has tended to investigate either particular question-types in organic chemistry assessments (often divorced from a classroom context) or has treated exam scores as simple outputs (Bhattacharyya & Bodner, 2005; Szu et al., 2011). But in actuality, exams are comprised of many different kinds of questions, all of which have different relationships with the curriculum. Without exploring the structure and context of exams and how these relate to student study and problem-solving behaviors we will not truly be able to understand what helps students to succeed in organic chemistry coursework. We may also not sufficiently understand the ways that exams are distinguishing students.

In this dissertation I seek to explore the relationship between study-behaviors, problem-solving strategies, exam design, and course context. I first begin in Chapter 2 by discussing the theoretical frameworks that form the foundation of this work, particularly the ideas of: exam structure and function, study behaviors, and problem-solving in organic chemistry. In Chapter 3 I then describe the methods used in data collection. Further detail on the methods used for data analysis are described in subsequent data analysis chapters. In Chapter 4 I describe the structure of the course in which this study is situated, Organic Chemistry 2, and the demographics of the student body included in the analysis.

Having situated this work both within the literature and within the real-world context of Organic Chemistry 2, the remainder of this dissertation focuses on answering four main research questions.

1) How are students' study behaviors associated with their overall exam performance in a second semester organic chemistry course (Organic Chemistry 2)?

2) How are students' study behaviors associated with their performance on different types of organic chemistry questions (specifically predict-the-product and mechanism problems)?

3) How are students' study behaviors associated with problem-type and with exercise-type questions (specifically within the sub-category of the predict-product questions)?

4) How do students' choices when problem-solving influence student performance on exercise-type and problem-type questions?

Chapter 5 focuses on question 1 and begins by exploring how students' study behaviors in Organic Chemistry 2 are associated with overall exam performance. It utilizes hierarchical linear modeling (HLM) and subsequent use of subject-mean centered covariates to answer research question 1. After identifying study behaviors significantly correlated with student performance, the impact of deviation from students' average usage of a resource on their performance on exams is explored. This chapter also serves as a source of comparison for subsequent chapters, allowing for deeper exploration on whether different types of questions are correlated with the same study behaviors as overall exam performance.

Chapter 6 explores how student study behavior is associated with different types of problems in order to answer research question 2. It specifically looks at two of the most common types of organic chemistry problems: Mechanism and Predict-the-Product problems to explore whether different types of problems are associated with different types of study behaviors. HLM and then subsequent use of subject-mean centered covariates are used to explore whether different types of questions associated with different study behaviors and whether study behaviors that are associated with overall performance are also associated with specific question types.

Chapter 7 describes the use of multiple linear regression to determine how students' study behaviors are associated with problem-type and exercise-type questions to answer research question 3. It explores two predict-the-product questions from the first midterm with different relationships with the curriculum and explores whether problem-type and exercise-type questions correlated with different study behaviors.

Finally, Chapter 8 describes a deeper exploration of student problem-solving on exam questions through thematic analysis of student think-aloud interviews to answer research question 4. This investigation sought to determine how use of particular problem-solving strategies might help or hinder students on problem-type and exercise-type questions. The tensions that exist between various problem-solving strategies when students solve problems was explored particularly focusing on the decisions students make when problem-solving and the forms of reasoning they consider convincing.

Chapter 2: Theoretical Framework

Examining Exams

Education researchers have been having conversations about the design and use of summative exams for many years and have explored this topic from many different lenses. One large subfield has discussed differences between formative and summative assessments (Bransford, Brown, & Cocking, 2000). This work has normally advocated for the increased use of formative assessments that provide students with more consistent and low-stakes feedback over the use of summative assessments. Others in the field have talked about the design of effective summative assessments. Some of this work has focused on whether assessments are designed around central concepts and test understanding over rote memorization (Stern & Ahlgren, 2002). Others have considered the affordances and allowances of the item design. For example, Linn et al.'s work which found that multiple choice questions are less effective because they mostly test recall over conceptual understanding (Linn, Lee, Tinker, Husic, & Chiu, 2006).

These conversations also continue in the field of chemistry education. The American Chemical Society, a professional scientific society supported chemists in the United States, has developed a series of national standardized tests for use at the university level (Holme & Murphy, 2012; Raker, Holme, & Murphy, 2013). The creation of these exams was motivated by a desire for standardized exams in chemistry built off of valid educational principles. These exams have developed from a set of ten anchoring chemistry concepts that extend across the subfields of chemistry (Murphy, Holme, Zenisky, Caruthers, & Knaus, 2012). However, most of the work on the creation of these exams has focused on anchoring them into a solid and well-reasoned conceptual foundation and has not considered the structure of the questions used. This work has also, by the very nature of being focused on standardized tests, not considered the relationship of the questions to the curriculum and course context.

Some studies, however, have considered the role of item design and relationship to the curriculum. Holme et al.'s work, for example, discussed alternatives to common assessment strategies so that they expanded beyond conceptual understanding to better support curriculum reform (Holme et al., 2010). Several studies have investigated the relationship between scores on different question types and overall exam performance (Austin et al., 2015; Webber & Flynn, 2018). These studies have concluded that certain question types are more strongly correlated with overall exam outcomes than others. While it is difficult to extricate whether these results indicate more about the nature of the exams or the strength of particular questions types, it does indicate that types of questions are functioning differently and that it is worth considering differences in question types within the field of organic chemistry.

Several studies have gone even further and treated different types of question as unique outcomes which might have differing relationships with factors such as student characteristics and curriculum. For example, Pribyl and Bodner's work showed that students with better spatial abilities received higher scores on chemistry questions, but only when the questions required problem-solving. (Pribyl & Bodner, 1987). Spatial ability did not predict performance on questions that could be treated as exercises. Zoller et al.'s work has distinguished between algorithmic problems and conceptual problems and examined student-problem solving in both contexts (Zoller, Dori, & Lubezky, 2002; Zoller, Lubezky, Nakhleh, Tessier, & Dori, 1995). Webber and Flynn also recently considered students solving of familiar and unfamiliar problems, and in particular how a new organic chemistry curriculum influenced student performance on each type of questions (Webber & Flynn, 2018). The works is beginning to consider the differences between types questions. What remains missing, however, are clearer distinctions in how these categories are defined as many seem to group together categories of questions with the questions' relationships to the curriculum.

Overall chemistry education researchers have considered the importance of exams being rooted in a well-defined conceptual framework, but are only beginning to more deeply consider other characteristics of exam design, particularly in the field of organic chemistry. This means that many studies use exam scores as simple outcomes variables and do not consider the make-up of these exams in their analyses(Paulson, 1999; Szu et al., 2011). While there is certainly work that acknowledges differences in the questions that make-up exams these categories are often vague and do not distinguish between categories of questions (for example multiple choice versus free response) and relationship to the curriculum. There has also been very little work that has considered how factors that influence student performance might interact with these different components of exams, with some exceptions for work exploring students' spatial abilities(Pribyl & Bodner, 1987). In this dissertation, I explore two potential aspects of assessment design: the structure of the questions being used and the nature of the questions' relationship to the curriculum. I am also interested in how students' study behaviors and problem-solving strategies may differently impact student performance on these different components of exams.

Categories of Questions

The first area of interest in this dissertation is the structure of the questions being used in assessments; this will be referred to as the category of the question throughout the dissertation. Well known examples of this are multiple-choice and free-response questions which are each their own unique category of question with affordances and allowances (Linn et al., 2006). Different disciplines however often have categories of questions that are unique to their field. In mathematics class, for example, a student might be asked to develop a geometric proof(Schoenfeld, 1988), a type of question that would be nonsensical in an English class. Organic Chemistry also has its own unique categories of questions. Much of organic chemistry relies on visual representations of molecules, and these representations are used to explain and predict how chemical processes occur. This means that organic chemistry utilizes unique categories of questions that ask students to utilize these representations to explain and predict reactivity. Two very common categories of organic chemistry questions are: Mechanism questions and Predict-the-product (PTP) questions. Mechanism questions ask students to represent the movement of electrons and atoms over the course of a reaction that lead to a product. Predict-the-product questions, in contrast, ask students to predict the outcome of a reaction given starting materials. For more details on these types of questions see Chapter 3 and Chapter 6. This dissertation explores whether students need to prepare in similar or different ways for each of these types of questions.

Problem-type and Exercise-type Questions

Many researchers have considered the relationship of questions to the curriculum, though these relationships have often been loosely defined and researchers have not usually distinguished between the questions' relationship to the curriculum and the structure of the question. While many instructors may have an intuitive understanding of the relationship a problem has to their curriculum, this is often couched under the idea of a problem's difficulty rather than more explicitly the extent to which aspects of the problem have been covered in the course. One recent example of work utilizing this approach, however, is Webber and Flynn's work (2018), which discusses familiar and unfamiliar problems within synthesis questions.

In order to explicitly describe the relationship of questions to the curriculum I borrow from the organic chemistry problem-solving literature which often emphasize the distinction between problems and exercises (Bodner & Herron, 2002). The field has often borrowed Hayes's (1980) definition of a problem which states that, "Whenever there is a gap between where you are now and where you want to be, and you don't know how to find a way to cross that gap, you have a problem"(pg. xii). In contrast, an exercise is a situation in which an individual can more easily identify a path to a solution. Bodner and Domin emphasize that the difference between a problem and an exercise is not difficulty, but the personal familiarity that a problem-solver has with the question (Bodner & Domin, 2000). There are questions that might be problems for students, but exercises for instructors who are more familiar with the subject matter.

While it is impossible to determine whether or not a given question would be a problem or an exercise to an individual without probing their thinking, one can determine whether a question is a problem or an exercise in reference to a given curriculum. Through analysis of the material presented in a course through lectures, practice problems, and any other resources one can measure whether a given question closely or distantly aligns with the curriculum. If a question is highly similar to examples seen previously in a course, either in lecture or on problem sets, a student has more opportunities to develop rote or algorithmic strategies characteristic of exercises even if the problem itself is not inherently algorithmic. In contrast if a question involves synthesis of various ideas presented in the curriculum or asks students to apply ideas to a new context then it functions as a problem. Questions can therefore be described as **exercise-type** questions, which are very similar to examples and material previously seen in the course, or as **problem-type** questions, which involve synthesis of multiple ideas or applying concepts in a new context. Problem-type questions would also include questions that were based on material that was only briefly covered in the course. These definitions don't imply that all students will view these questions as problems or exercises; instead they describe the relationship the question has to the curriculum and whether or not students are likely to view them as problems or exercises. By determining this relationship, we can better determine what student strategies, either in how they study or problem-solve, allow them to bridge the gap between the curriculum and the question.

Study Behaviors

Chapters 4-6 of this dissertation focus on student study behaviors and how they are related to student performance in Organic Chemistry 2. Most of the work investigating student study behaviors and motivations have been based on Pintrich's Self-Regulated learning model (SRL) (Pintrich, 2004). This is a top-down approach based on psychology theory that assumes that knowledge is constructed by students, that students have the potential to control, or regulate their cognition, motivation, behavior, and some aspects of their environment, that student can and do set academic goals, and that self-regulatory activities mediate between the person, context, and eventual achievement. While this work has had important insights on the influence of study behaviors on student performance, the top-down approach mean that it does not always capture the broad array of study behaviors that students engage in.

Interesting results in this field have been observed by researchers utilizing alternative frameworks and constructions based on a more bottom-up approach which allows them to develop findings based on study strategies/approaches unique to college chemistry coursework. Much of this work has been limited by small sample sizes which have allowed for detailed analysis of small populations through interviews and journaling but have not allowed for more detailed quantitative analysis (Szu et al., 2011). Other work, like that of Sinapuelas and Stacy (2015), have been able to the explore the question of student study behaviors in both a bottom-up and quantitative way but has done so within a freshman general chemistry course, which is very different from organic chemistry in both the material and types of question used in assessments. They also found that their quantitative instruments did not sufficiently capture differences in student study strategies as many students replied that they used the majority of strategies listed and instead better measured differences in their approaches.

Sinapuelas and Stacy's work is derived from a larger field of work exploring student approaches to learning (SAL) in a variety of fields. This field has been developed from Marton and Saljo's work on student approaches to learning in a reading assignment (Marton & Saljo, 1976). In this work they first coined the term approaches to learning and the idea of deep-level and surface-level learning. Surface-level learning refers to students focus on memorization or collecting disconnected facts, whereas deep-level learning refers to learning in which students focus on underlying concepts and truly understanding the material. This work usually also is closely tied to student motivation, with surface-level approaches tried to extrinsic motivation and deep-level approaches tried to intrinsic motivation (Entwistle & McCune, 2004).

In this dissertation I utilize a new theoretical framework, which I refer to **Student Study Behaviors**. I define study behaviors as choices students make in how to engage with the resources provided to them within a given course. This concept is related to work conducted on approaches to learning, particularly that by Sinapuelas and Stacy (2015). However, while Sinapuelas and Stacy utilized, an interview approach, this dissertation utilizes a quantitative survey instrument developed around the resources available to students in Organic Chemistry 2 (See Chapter 3 for more information on the development of the instrument).

The Student Study Behaviors framework is based on the idea that study behaviors are not simply representative of underlying learning approaches, but are important to understanding how students utilize the resources available to them within a specific context. This in-depth measurement allows for differentiation on what behaviors would be most useful to students in receiving high scores on course assessments (or components of course assessments). This framework is a useful lens for exploring the relationship between exam design, study behaviors, and student performance. Information on the effectiveness of given study behaviors in helping students prepare for exams (or other types of assessment) can also be more easily translated into guidance for instructors and students.

By focusing on study behaviors instead of student approaches to learning I am also able to disconnect students' behaviors from students' motivations in the course. While much of the education literature has shown that extrinsic motivation is correlated with higher performance, prerequisite courses such as organic chemistry are often taken by students who are not particularly interested in the material but still desire to do well. Organic Chemistry 2, as a class for non-majors, is notable in that the majority of students are extrinsically motivated. This means that measures centered on student motivation such as SAL are poorly suited to this population. Use of an alternative approach such as study behaviors, allows this study to provide information on what behaviors lead to better outcomes, even in situations in which students are not intrinsically motivated to perform well in the course.

Problem Solving Strategies

The field of chemistry education has explored the problem-solving strategies that students engage in and whether they are effective in helping students solve problems more accurately. This research has been mostly consisted of smaller qualitative studies, usually asking students to participate in think-aloud interviews in which they solve example problems. Researchers in the field of organic chemistry problem-solving tend to explore students' work in mechanism questions and, in rarer cases, synthesis questions (Bodé & Flynn, 2016). Some results in the field have indicated that skills like representational competence (Bodner & Domin, 2000) are important for student problem-solving. Others have identified more specific strategies like drawing mechanisms and identifying reactive sites (Bodé & Flynn, 2016).

An overarching approach throughout the field is comparing students who solve problems algorithmically to those that solve problems conceptually. This comparison has been particularly emphasized in research examining problem solving in general chemistry (a prerequisite course to organic chemistry). Nurrenbern and Pickering, for example, identified that there is a difference between problem solving and conceptual understanding, and that it was problematic to assume that they were interchangeable (Nurrenbern & Pickering, 1987). Organic chemistry education researchers have investigated the related area of reasoning strategies (Christian & Talanquer, 2012; Kraft et al., 2010). Their work has compared rules-based reasoning, case-based reasoning, model-based reasoning, and, in some work, symbol-based reasoning. Reasoning strategies like rules-based reasoning and symbol-based reasoning, align closely with algorithmic thinking, representing types of thinking in which students utilize rules or manipulate drawings of chemicals with no real understanding of what the drawings represent. Case-based and model-based reasoning, in contrast, more closely resemble conceptual skills, because they require students to either recognize patterns and trends based on previous examples or for students to understand the chemical models underlying the problem.

While work on organic chemistry problem solving has acknowledged that students utilize multiple reasoning strategies while they problem solve, research has tended to focus on the general strategy that a student utilizes while solving a given question (Christian & Talanquer, 2012; Kraft et al., 2010). In this dissertation, however, I was interested in the reasoning that students use throughout their process of solving a question and more specifically students' decision-making process as they solve organic chemistry questions. There has been a lot of work in the field of education on the nature of problem solving, some that posits that problem solving can be seen as a fairly organized process (Polya, 1945) and some that describes problem solving as disorganized and cyclical (Bodner & Domin, 2000).

A helpful framework for considering student decision-making and reasoning within a while they problem-solve is the idea of a problem space. Problem space, also known as a search space is the idea that each problem consists of a series of possible answers and that the problem-solver must search within them for the possible answer (Klahr & Dunbar, 1988; Anderson, 2009). Problem-solvers must seek operations (or strategies) that allow them to navigate this space and will often seek to constrain the problem-space through information given in the problem.

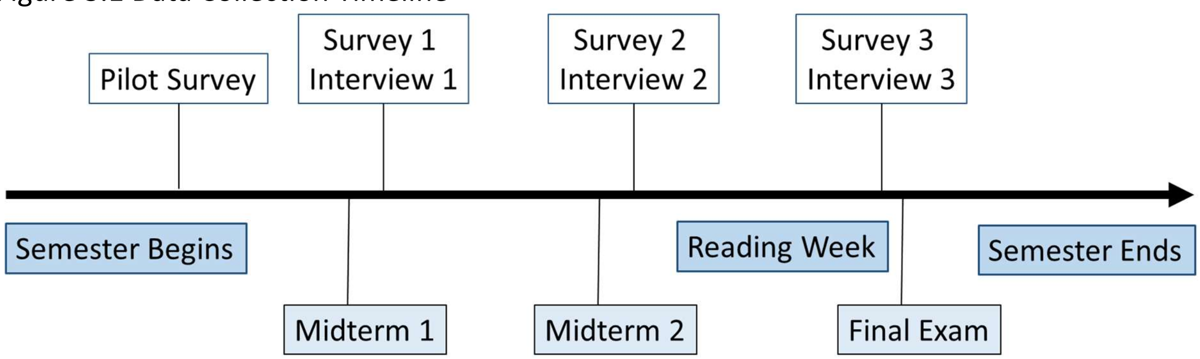
Since researchers in the field of organic chemistry have not explored the details of student problem-solving there is less information on the phases and reasoning strategies that make up organic chemistry problem-solving. Therefore, for the work in Chapter 8 analyzing student problem solving, reasoning strategies and problems-solving phases were identified via thematic coding of student think-alouds (See Chapter 8 for more information on the methods used and reasoning strategies identified).

Chapter 3: Methodology

Introduction

For this dissertation I utilized an integrated mixed methods approach with quantitative and qualitative data collected simultaneously. Mixed methods were necessary in this study since the research questions could not be answered with exclusively quantitative or qualitative methods (Creswell & Plano Clark, 2011; Johnson & Onwuegbuzie, 2004). In particular, use of student interviews allowed for deeper analysis of student problem solving, and observations were critical for contextualizing exams and study behaviors in a course context. The goal of this data analysis was to explore how study behaviors and problem-solving strategies influenced student performance in [Organic Chemistry 2], a second semester organic chemistry course at [University of the West Coast]. This study utilized the simultaneous collection of six sources of data: (1) Surveys, (2) Student Interviews, (3) Classroom Observations, (4) Course materials (Instructor notes, homework assignments, etc.), (5) Student exam scores and final grades, (6) Student work (Responses to exam questions). Approval of the Internal Review Board (IRB) of the university was received in 2018.

Figure 3.1 Data Collection Timeline



Surveys

In order to measure student study behaviors in organic chemistry a survey instrument was developed through a multi-step design process. First a list of resources available to students in Organic Chemistry 2 and how they could be used was made. This list was developed based on my own experiences previously as a chemistry graduate instructor at University of the West Coast, through conversations with chemistry instructors at the University of the West Coast (particularly the instructor who was to teach Organic Chemistry 2 in the Spring 2018 semester), and through information in the literature on student approaches to learning and problem solving in organic chemistry (Kraft et al., 2010; Sinapuelas & Stacy, 2015). Once the list had been developed, the resources were divided into five main categories: In-class study behaviors, use of additional instructional resources, use of written/online resources, use of practice problems, and engagement in group work. Questions were then developed to probe how students chose to use these categories of resources. In general, for each resource there tended to be opportunities to use it more actively (to engage with the content and participate) and more passively (to listen and accept information). For example, in taking notes in class students could either copy down what the instructor said or wrote verbatim (passively) or translate ideas into their own words and write down their own thoughts and ideas (active).

Questions were also developed exploring student demographics, academic backgrounds, and asking students to reflect on their studying and problem solving.

The preliminary instrument was then evaluated by the course instructor, as well as by other chemistry instructors and education researchers. It was also evaluated by a group of six undergraduate researchers, five of whom had previously taken Organic Chemistry 2. The remaining undergraduate was a transfer student and had taken organic chemistry at another institution. Feedback was collected on whether any resources or usages were missing and general interpretability of the survey. Finally, a pilot survey was sent out to students in Organic Chemistry 2 during the third week of class. 59% of students completed the pilot survey (N = 320). This survey was used to collect demographic data, recruit interview participants, and also to determine any potential missing study behaviors. Most questions were designed with an "Other (please describe)" response which allowed students to describe any missing resources or usages. Classroom and review session observations were also used to identify any other resources or study behaviors, particularly those present in the classroom. The survey was then edited to incorporate student responses

Subsequent surveys using the final instrument were conducted after each midterm, and during Exam week (See Figure 3.1). Students completed the first and second survey the week after each midterm. The final survey however was given after Reading week (a week provided to students to study and prepare for final exams) in order to capture student behaviors during this time but was given during Exam week to maximize student participation. This meant that some students completed it prior to the final exam and some after. In accordance with our IRB protocol, all students were given the opportunity to participate in the survey or an alternative essay assignment and were given two-point extra credit in the course for completion of either the survey or the alternative assignment. Students included in this project also gave permission for their grades and course materials to be used for research purposes.

The surveys asked students to reflect back to the previous week on their study behaviors. Students were particularly asked to reflect on their participation in additional forms of instruction (review sessions, tutoring, etc.), their usage of written resources (notes, textbooks, etc.), their usage of practice problems, and their participation in group work. There were also reflection questions about how they studied for the exams and how they would approach solving example PTP questions. The pilot and last surveys also asked questions about their academic and personal backgrounds (See Appendix A for an example survey text). The surveys had identical items except that the final survey included questions about academic and personal backgrounds and reflective problem-solving questions were modified to include problems from recent quizzes or exams.

Variables

A subset of survey questions was used to answer the first three research questions in Chapters 5-7. Variables for analysis were created from survey responses to these items; details on their construction is included below and table of frequency counts for the different study behaviors can be found in Appendix D. While the survey asked students about their usage of resources provided both by a variety of different sources this analysis focuses exclusively on resources provided within the course. Chapters 5-7 all explore the same resources and associated study behaviors and therefore share many of the same independent variables. There are, however, differences in the subject mean-centered covariates and outcome variables used, so these will be described later in the dissertation in their relevant chapters. Also note that Chapters 5 and 6 discuss results analyzing data from all three time points with hierarchical linear modeling. Chapter 7 however describes the analysis of data from the first exam and survey with multiple linear regression. While each of the variables is described as either a level 1 or a level 2 covariate this distinction is only relevant in chapters 5 and 6. Level 1 covariate refers to variables that are time-variant and Level 2 covariates to variables that are time-invariant. The occasion variable was also not used in Chapter 7 since only one occasion was being analyzed.

Control Variables

occasion: A categorical level 1 variable corresponding to the time point at which the survey and exam scores were collected. The first survey and the first midterm correspond to Occasion 1, the second survey and the second midterm correspond to Occasion 2, and the third survey and final exam correspond to Occasion 3. Occasion 1 was treated as the reference category and dummy variables were created for the subsequent occasions: "Occasion 2" and "Occasion 3." This variable was introduced to control for differences in difficulty between the exams.

Study hours

study_hours: A categorical level 1 variable corresponding to the number of hours the student studied the week before each exam. Students were asked to indicate how many hours they had worked including readings and problem sets but not including work from a related laboratory course. This variable was used to control students' time on task prior to each exam, though results from the third survey may be lower than reality since interviews indicated that many students chose to study the week of exams rather than the week before. While the initial question included options for "<1 hour" and "1-4 hours" these two categories were collapsed due to low response rates for the "<4 hour" category. Final categories consisted of: "<4 hours" which corresponded to 0, "4-7 hours" which corresponded to 1, "7-10 hours" which corresponded to 2, and "10+ hours" which corresponded to 3. "<4 hours" was treated as the reference category and dummy variables were created for the other three categories.

In-class study behaviors

Since Organic Chemistry 2 utilized an active learning pedagogy, students had additional resources in class that they could utilize. Four main types of in-class study behaviors were identified: Working-on and editing of in-class worksheets, collaboration with peers, actively taking notes, and engaging with the instructors. The final category was found to be a relatively uncommon behavior and therefore not included in this analysis. Since taking notes and working on worksheets was found to be a common behavior, active use of each of these resources (in which students engage with the material) was compared to more passive usages (in which students simply copy down information provided by the instructor).

In the survey students were asked to reflect on their in-class study behaviors the week before the exam. This was considered to be representative of students' normal behavior during lecture. These measurements may not completely reflect standard student behavior since student attendance and engagement was often lower the week before an exam because the material presented that week was not included in the midterms. Also, in the final week of classes students are often distracted by end-of-semester assignments and examinations. However, I decided to have students reflect on the previous week instead of an average in order to reduce recall bias.

lecture_collaboration: This is a dichotomous Level-1 variable that corresponds to whether participants talked/collaborated with neighbors during lecture. Participants were asked to select all the study behaviors that they used during lecture in the previous week. This variable was created by assigning a value of 1 to participants who selected that they "Collaborated/Talked with neighbors", 0 to participants who answered the question but did not select that they "Collaborated/Talked with neighbors", and a missing value to participants who did not answer the question.

notes_ownwords: This is a dichotomous Level-1 variable that corresponds to whether participants engaged in active note-taking. This variable was created by assigning a value of 1 to participants that indicated that they had engaged in active notetaking ("Translated what the instructor said into my own words and/or Wrote down my own thoughts and insights") and a value of 0 to participants that indicated that they had only engaged in passive notetaking ("Wrote down what the instructor said/wrote verbatim").

worksheet_revision: This is a dichotomous Level-1 variable that corresponds to whether participants engaged actively with in-class worksheets. Participants were asked to select all the ways that they used the in-class problems/worksheets. This variable was created by assigning a value of 1 to participants who selected that they "Work on the problem, then revise the solution when the instructor goes over the problem" and a value of 0 to participants who answered the question, but did not select this response. A missing value was assigned to participants who did not answer this question.

Use of additional instructional resources

Students were provided three main forms of additional instruction in the class: Weekly review sessions, Office hours (Instructor and GSI), and Piazza message boards. Other forms of additional instruction were available to students through the university (notably peer-led study groups and tutoring) but these are not affiliated with the class so were not included in this analysis. There is also some variation in how students could use office hours and review session which is not reflected in this analysis. Because student usage of each of these resources was relatively low, I decided not to explore further differentiation (See Appendix D).

review_session: This is a dichotomous Level-1 variable that corresponds to whether or not participants attended a review session the week before the exam. Participants were asked to select all the additional instructional resources that they utilized the week before the exam A value of 1 corresponds to the participants who selected "Review sessions held by head GSI" a value of 0 corresponds to students who answered the question but did not select "Review sessions held by head GSI".

office_hours: This is a dichotomous Level-1 variable that corresponds to whether or not participants attended office hours the week before the exam. Participants were asked to select all the additional instructional resources that they utilized the week before the exam. A value of 1 corresponds to the participants who selected either "GSI office hours" or "Instructor office hours", a value of 0 corresponds to students who answered the question but did not select either response. Instructor and GSI office hours were combined because they are structurally quite similar and separately response rates were too low for quantitative analysis.

piazza_active: This a categorical Level-1 variable that corresponds to how participants utilized piazza message boards the week before the exam. Participants were asked to select all the ways that they utilized the piazza message board. A value of 0 corresponds to participants who indicated that they did not utilize the piazza message, a value of 1 corresponds to participants who passively used the piazza message board (i.e., those who indicated that they only "Read responses to the questions of others"), and a value of 2 corresponds to participants who actively used the piazza message board (i.e,. those who responded "I asked questions", "I had my questions answered", or "I answered questions posted by others"). Participants who did not use the piazza message boards were used as a reference category and dummy variables were created for the other two categories.

Use of written or online resources

Students were asked to reflect on how much they used various online or written resources (see Appendix A for more details) and almost all students reported using at least one of these resources the week before their exams. Since use of these resources was common, I chose to explore how students utilized this category of resources rather than which individual resources they used. Students were therefore asked which of the following activities they predominantly did (and instructed to only select their top 3):

- Memorize key words, concepts, molecules, and reactions
- Review problem types that might appear on a quiz/exam
- Test my understanding of the material
- Build connections among ideas
- Build connections between the material and the real world
- Identify underlying concepts
- Review problem solving approaches

These categories of behaviors were developed from the literature and conversations with instructors and students. The category of "Build connections between the material and the real world" was dropped for this analysis because this type of resource was rarely chosen. Responses were also dropped for participants who indicated that they had only used resources unaffiliated with the course (specifically "Read other textbooks or wikis" and "Watched related videos"). Students needs to have used at least one resource provided by the course (specifically "Read the textbook", "Reviewed lecture notes", Reviewed midterms/quizzes", or "Reviewed in-class problems/worksheets").

All variables constructed in this section were dichotomous level-1 covariates corresponding to participants that indicated that they predominantly did a given activity when they reviewed written or online resources (See Table 3.1). Students who selected that they did a given activity were assigned a value of 1. A value of 0 and indicates that students answered the question but did not select that activity.

Table 3.1
Possible Responses for Use of Online/Written Resources Variables

Variable	0	1
review_memorize	Predominantly memorized key words, concepts, molecules, and reactions	Did not predominantly memorize key words, concepts, molecules, and reactions
review_Ptypes	Predominantly reviewed problem types that might appear on a quiz/exam	Did not predominantly review problem types that might appear on a quiz/exam
review_testunderstanding	Predominantly tested my understanding of the material	Did not predominantly test my understanding of the material
review_connectideas	Predominantly built connections among ideas	Did not predominantly build connections among ideas
review_concepts	Predominantly identified underlying concepts	Did not predominantly identify underlying concepts
review_Psolving	Predominantly reviewed problem-solving approaches	Did not predominantly review problem-solving approaches

Use of practice problems

Students were asked to reflect on how much they used various sources of practice problems (see Appendix A for more details) and almost all students reported using at least one of these resources the week before their exams. Since use of practice problems was common, I chose to explore how students utilized this category of resources rather than which individual resources they used. The same activities were identified for working on practice problems as for reviewing written and online resources. Students were therefore asked which of the following activities they predominantly did while using practice problems (and instructed to only select their top 3):

- Memorize key words, concepts, molecules, and reactions
- Review problem types that might appear on a quiz/exam
- Test my understanding of the material
- Build connections among ideas
- Build connections between the material and the real world

- Identify underlying concepts
- Review problem solving approaches

These categories of behaviors were developed from the literature and conversations with instructors and students. The category of "Build connections between the material and the real world" was dropped for this analysis because response rates for it were very low. Responses were also dropped for participants who indicated that they had only used resources unaffiliated with the course (specifically "Online problems (unaffiliated with the class)" and "SLC mock midterm"). Students needs to have used at least one resource provided by the course (specifically "Problem sets", "Practice exams", or "Textbook problems").

All variables constructed in this section were dichotomous level-1 covariates corresponding to participants that indicated that they students who predominantly did a given activity when working on practice problems (See Table 3.2). Students who selected that they did a given activity while working on practice problems were assigned a value of 1. A value of 0 and indicates that students answered the question but did not select that category.

Table 3.2

Possible Responses for Use of Practice Problems Variables

Variable	0	1
problem_memorize	Predominantly memorized key words, concepts, molecules, and reactions	Did not predominantly memorize key words, concepts, molecules, and reactions
problem_Ptypes	Predominantly reviewed problem types that might appear on a quiz/exam	Did not predominantly review problem types that might appear on a quiz/exam
problem_testunderstanding	Predominantly tested my understanding of the material	Did not predominantly test my understanding of the material
problem_connectideas	Predominantly built connections among ideas	Did not predominantly build connections among ideas
problem_concepts	Predominantly identified underlying concepts	Did not predominantly identify underlying concepts
problem_Psolving	Predominantly reviewed problem-solving approaches	Did not predominantly review problem-solving approaches

Use of group work

Participants were asked to reflect on how much time they spent on various types of group work the week before the exam. I chose to focus on students' behaviors in self-formed groups (ie. with a buddy, group of friends, or group in a residence hall or social organization). I chose to not include data on students who worked together in peer-led study groups since interviews indicated that students often did not work collaboratively in these study groups and many predominantly interacted with the peer-instructor.

grpstudy_active: This a categorical level-1 variable corresponding to students' level of participation in self-formed groups. Participants were asked to describe their behavior at their last group meeting. A value of 0 corresponds to participants who indicated that they did not engage in group work, a value of 1 corresponds participants who passively engaged in group work (ie. those who indicated “I mostly listened”), and a value of 2 corresponds to participants who actively engaged in group work (ie. those who responded “I mostly asked questions”, “I asked and answered questions” or “I mostly answered questions”). Participants who did not engage in group work were used as a reference category and dummy variables were created for the other two categories.

Student Exam Questions and Scores

Exam and quiz data were also collected throughout the semester. Student exams were graded using Gradescope, and a copy of the Organic Chemistry 2 site was made that only included the responses of participants in the study. Participant exams and quiz questions were used as a source of example problems for the surveys and think-alouds. Exam scores, as well as predict-the-product and mechanism question sub-scores, were also used as outcome variables measuring student performance for some of the research questions. All exam scores, and problem-type sub-scores were normalized by converting to percentage scores.

Student Interviews

Interviews were conducted at three time-points (after the 1st and 2nd midterms during Reading week and exam week). At the first- and second-time points students were interviewed within two weeks of each exam. For the final interview students were interviewed over the two weeks of Reading week and exam week. Some students completed the interview prior to the exam and some after the exam was completed. Because most students leave quickly after exams it was not possible to interview all the students after they completed their final.

Volunteers were recruited during the first pilot survey and asked to provide email addresses where they could be reached. From this sample of volunteers were selected participants who had completed Organic Chemistry 1 (the most common prerequisite course) for the first time in Fall 2017 and who provided their grade in the course (N = 75). Ensuring that all students had taken the same prerequisite organic chemistry course allowed me to control for academic background and experiences and to make better comparisons between students. These participants were then emailed and asked to complete a first-round interview. Students were not offered any compensation though snacks were provided during the interview. Twenty-two participants were recruited (see Table 3.3 for a break-down of demographics), though not all participants completed every round of interviews. Interviews were video recorded though only students’ hands and writing were captured to ensure anonymity. Interviews lasted on average between 30-45 minutes.

Table 3.3
Interview Participants

Participant	Gender	Ethnicity	Organic Chemistry 1 Grade	Organic Chemistry 2 Grade	Time 1	Time 2	Time 3
Tessa	Female	East Asian	A+	A+	Yes	Yes	Yes
Neil	Male	White/Caucasian	A+	A	Yes	Yes	Yes
Jakob	Male	Middle Eastern/North African	A+	A	Yes	Yes	No
Andrew	Male	White/Caucasian	A	A+	Yes	Yes	Yes
Andrea	Female	South Asian; White/Caucasian	A-	A	Yes	Yes	Yes
Rebecca	Female	White/Caucasian	A-	A-	Yes	Yes	Yes
Patricia	Female	East Asian	A-	A+	Yes	Yes	Yes
Monica	Female	White/Caucasian	A-	B+	Yes	No	No
Kelvin	Male	East Asian	A-	A-	Yes*	Yes	Yes
William	Male	East Asian	B+	B+	Yes	No	Yes
Katie	Female	East Asian	B+	B	Yes	Yes	Yes
Lulu	Female	East Asian; White/Caucasian	B+	A	Yes*	Yes	Yes
Sophia	Female	Mexican American/Chicano; White/Caucasian	B+	A-	Yes	Yes	Yes
Amanda	Female	White/Caucasian	B+	A	Yes	Yes	Yes
Mia	Female	East Asian	B+	A	Yes	No	No
Keira	Female	White/Caucasian	B+	B+	Yes	No	No
Nancy	Female	Filipino; White/Caucasian	B	A-	Yes	Yes	Yes
Olivia	Female	White/Caucasian	B-	C	Yes	Yes	No
Heather	Female	White/Caucasian	C	B	Yes*	Yes	Yes
Brianna	Female	Black/African American; White/Caucasian	C-	W	Yes	Yes	Yes
Samantha	Female	White/Caucasian	C-	C	Yes	Yes	Yes
Jessica	Female	East Asian	C-	C	Yes	No	No

*Indicates that incomplete audio was collected for the observation.

Interviews were conducted using a semi-structured interview protocol (See Appendix B for sample interview text). The first half of the interview consisted of questions asking students to reflect on the exam they had recently taken and how they had chosen to study for it. Follow-up and probing questions were asked to ensure that as many details about student study practices/approaches had been captured. The second half of the interview consisted of asking students to think-aloud on 2-4 PTP exam questions (Solomon, 1995). Some variation in number of questions occurred based on the length of the first half of the interview. Students were instructed to share what they were thinking and not to provide an explanation. They were also told that they could cease working on the problem whenever they wanted. After completing each problem students were then asked follow-up questions about their decisions and problem-solving. They were also asked to reflect on how their studying had helped prepare them for the question and what they could have done, if anything, to better prepare for the problem. In the final round of interviews, students were asked to reflect on questions that I developed based on material they had covered in class and problem sets since the second midterm. These questions were then reviewed by the instructor to ensure that the questions were consistent with the style and content of her assessments. Also, since many students were interviewed prior to the final exam, this also ensured that questions from the interview would not be overly similar to exam material.

This dissertation will focus on the first round of student interviews. Fifteen students successfully completed think-alouds in the first interviews and are used in this analysis. Interviews were first paraphrased by myself and a team of undergraduates. After initial paraphrasing the interviews were transcribed fully. To do this the interviews were converted into mp3 files and transcribed using Amazon Web Services Transcribe software. These initial transcripts were then edited using Inqscribe to produce the transcripts that were subsequently analyzed in Chapter 7. Transcripts not only included participants utterances but also descriptions of their actions as they thought-aloud.

Classroom Observations

Observations were conducted during every course lecture and during the weekly review sessions hosted by the head GSI. Two observers were present for the majority of observations (with some exceptions due to scheduling issues or emergencies). Notes were taken in one of two styles: Table and Annotated Chronology (see Appendix C). When possible both styles were used since the table allowed for greater organization of data while the annotated chronology afforded more detailed reporting. Observations were focused on the instructor/head GSI, the topics they covered, and pedagogical practice they used. The following areas were the main focus of the observations: Topics covered in class, Pedagogical Choices/Active Learning strategies used by the instructor, Out-of-class behaviors encouraged by the instructor, Problem-solving strategies used by the instructor, and Predict the product questions discussed in class. Immediately after each observation observers would have a debrief meeting to discuss the observation. They would then summarize their observations into a summative table. In addition to observation notes we also collected the instructors written notes, which she would post online after each lecture (the review session GSI did not post such notes). Together the observation notes and instructor notes provide a detailed report of what was said and written during the lectures and review sessions. The summative table provided information on overarching trends across the semester.

Chapter 4: Course Description and Student Demographics

Course Description

In order to explore the interplay between instructor pedagogy, student study/problem solving behavior, and student outcomes I chose to study a second semester organic chemistry course for non-majors: [Organic Chemistry 2] at a prestigious, research-intensive, public university on the west coast: [University of the West Coast]. A second semester organic chemistry course was selected over a first semester course for two main reasons: the instructor teaching the second semester course utilized active learning pedagogy, which allowed me to study students' use of in-class recourses instead of just out-of-class resources, and second semester organic chemistry tends to utilize more consistent question types in assessment than first semester organic chemistry courses, allowing for exploration of the relationship between question designs and student outcomes.

At the beginning of the semester, 538 students were enrolled in Organic Chemistry 2. Class enrollment data was collected during the second week of the semester since there is often fluctuation in course enrollment during the first week. Five hundred and nineteen students completed the course; the remaining nineteen either dropped the course or received an incomplete. As previously stated, Organic Chemistry 2 was taught utilizing active learning pedagogy, notably the instructor provided students with notes handouts that they worked on in class. Course lectures occurred twice a week and lasted 80 minutes each; students could either attend in the morning or in the afternoon. Students were encouraged to fill out the handouts, complete problems in class, talk to their neighbors, and ask questions of the instructor or graduate student instructors (GSIs). Organic Chemistry 2 covered the following topics: Acids & Bases, Resonance, Aromaticity, Allylic Systems, Carbonyls, Protecting groups, Hydrates & acetals, Imines, Acyl substitution, Reduction/oxidation of carbonyls, Wittig reaction, Enol/Enolate chemistry (ie. Aldol and Claisen condensations), Electrophilic Aromatic Substitution, Nucleophilic Aromatic Substitution, and Diels Alder reactions. Throughout the semester the instructor also heavily emphasized biological connections to the reactions and concepts learned.

The course utilized a flat grading scale and student scores were based on: weekly quizzes, participation in iClicker questions, two midterms, a final exam, and an extra credit assignment (assigned after the first exam resulted in lower scores than the instructor anticipated). Students were also provided with problem sets after each lecture. They were strongly encouraged to complete the problem sets, but they were not graded for points. Problem sets and assessments tended to be structured around five types of problems:

- Nomenclature
    - Nomenclature questions ask students to either provide the name for a given chemical or to draw a chemical structure given a chemical name.

- Mechanism
    - In mechanism questions students were given a chemical reaction and asked to use arrow-pushing notation to indicate how electrons move through-out the reaction to result in the given product.
- Biosynthesis
    - Biosynthesis question were a subset of mechanism questions in which students were provided with information about a reaction occurring within a biological system and asked to draw the mechanism of this reaction. This type of question was used by the instructor to further emphasize the biological applications of organic chemistry.
- Explanation
    - Explanation questions asked students to explain why a chemical phenomenon had occurred.
- Predict-the-Product/Synthesis
    - Predict-the-product questions provide students with starting materials and ask them to predict the products of the reaction. Synthesis questions are a related type of question, but instead of the product, students are given starting materials and a product and asked to either provide the necessary reagents or relevant intermediate.

Students also had opportunities to utilize a course-wide piazza message board, instructor and GSI office hours, and weekly review sessions run by the head GSI. Outside of class many students utilize resources provided by the Student Learning Center (SLC), namely tutoring, study groups, and mock midterms. All SLC resources are run by peer tutors. Students also reported using online resources like online problems, wikis, and videos.

Participant Demographics

Of the 538 students in the course, 91% completed at least one survey (N = 488) and 54% completed all three (N = 289). Four hundred and sixty-nine students were included in the quantitative analyses conducted in chapter 5 and 6 and 357 students were included in the quantitative analysis conducted in chapter 7. The study population was compared to institutional data on course wide grade data and seems fairly representative in terms of grade distributions (See Figures 4.1 and 4.2), though there does appear to be slight overrepresentation at the higher grade levels, particularly students who received an A- and underrepresentation at the lower grade levels, particularly students who received a C. Also note that due to fairly early drop deadlines many students who received an F or a NP may have been those who effectively chose to drop the course.

Figure 4.1 Comparison of Course Wide and Population Grade Distributions

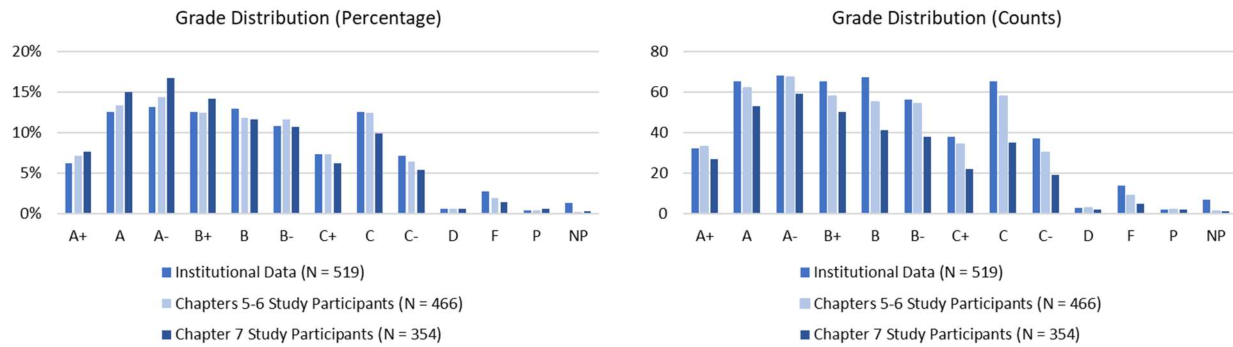
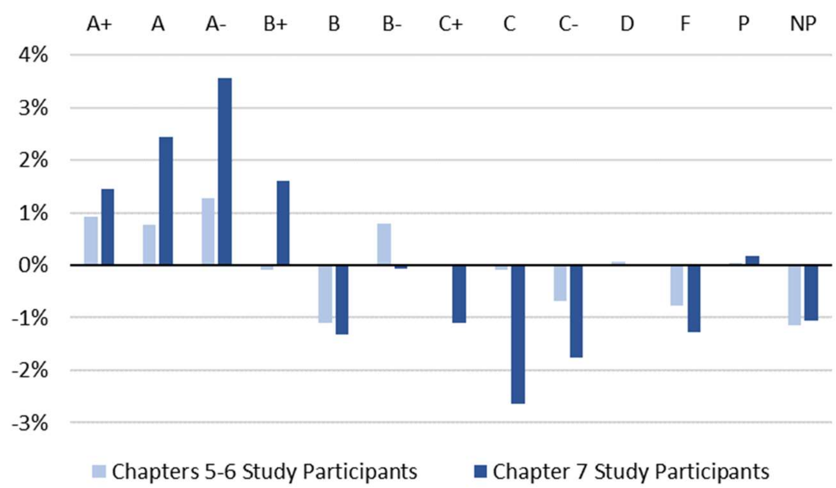


Figure 4.2 Differences between Total Percentage Grade Distribution and Population Grade Distributions



Institutional data on student demographics was not collected but future work will use institutional data to compare the study population to the whole course. However, within the Chapter 5-6 population 69% of students identified as female (*N* = 298) and 31% as male (*N* = 139). The Chapter 7 population was very similar with 69% of students identifying as female (*N* = 229) and 31% identifying as male (*N* = 103). The low percentage of male students suggests that the study sample may not be representative of the class as a whole and that male students may have been less likely to fill out the survey. Students were predominantly White/Caucasian or East Asian (See Figure 4.4) and the majority had at least one parent/guardian with a graduate degree (Figure 4.3).

Figure 4.3 Parents' Education

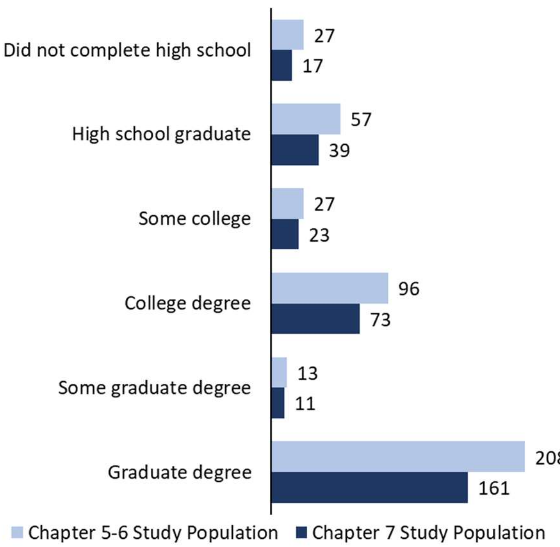
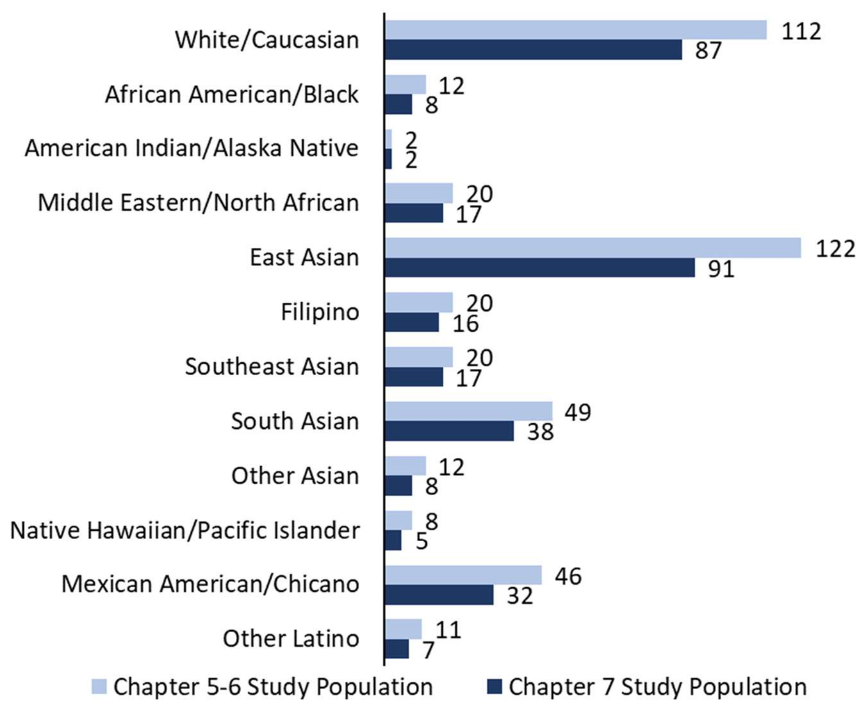
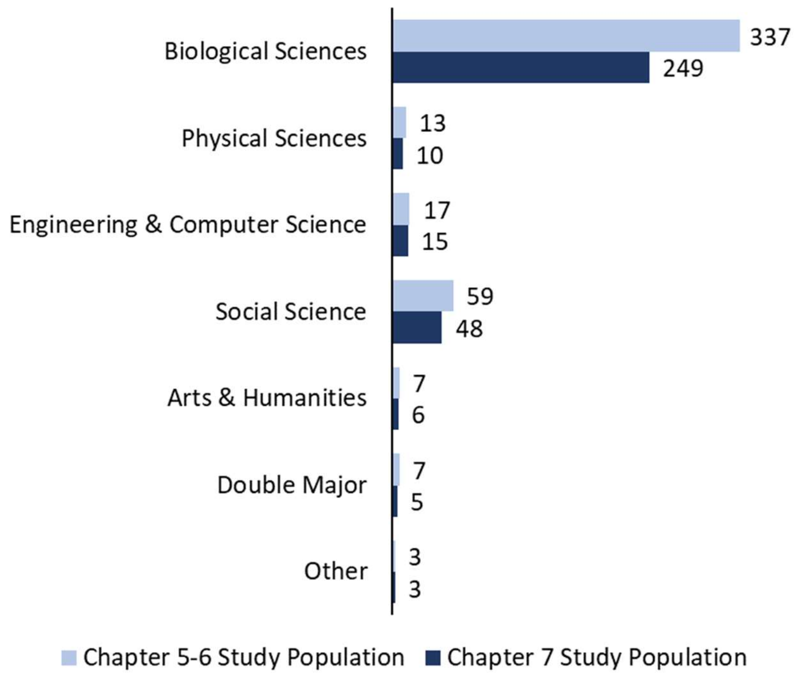


Figure 4.4 Student Ethnicity



Study participants were also predominantly majoring in Biological Sciences or in Social Sciences (See Figure 4.5). For the Chapter 5-6 Study Population 70% were in the premedical track (*N* = 310). The Chapter Study Population was similar with 73% of participants on the premed track (*N* =245). Students in the Chapter 5-6 study population took an average of 15.29 units (*SD* = 2.32, *Min* = 9, *Max* = 30, *N* = 441) in the Spring 2018 semester. The Chapter 7 population was almost identical with students taking an average of 15.24 units (*SD* = 2.36, *Min* = 9, *Max* = 30, *N* = 335). 41% of students in the Chapter 5-6 Study population worked (*N* = 180), and those students worked an average of 11.12 hours (*SD* = 7.35, *Min* = 1, *Max* = 50). 41% of students in the Chapter 7 Study population worked (*N* = 137), and those students worked an average of 11.13 hours (*SD* = 7.20, *Min* = 1, *Max* = 50).

Figure 4.5 Study Population Majors



When asked in a free-response question what their motivations were for the course 95% of the Chapter 5-6 population indicated that at least one of their reasons was the that the class was a requirement (*n* = 419) many indicating that it was a premed or major prerequisite and 85% indicated it was their only reason for taking the course (*n* = 373). 11% indicated that they took the course out of interest or enjoyment (*n* = 48) and 4% because they thought the material might be useful to them in the future (*n* = 19). The Chapter 7 study population was similar with 94% of the population indicating that at least one of the reasons they took the class was that the class was a requirement (*n* = 314) and 83% indicated that this was their only reason (*n* = 277). Only 12% of students indicated that they took the course out of interest or enjoyment (n = 41) and only 5% indicated that they thought the material might be useful to them in the future. This indicates that students in Organic Chemistry 2 are predominantly extrinsically motivated.

Chapter 5: Study Behaviors and Exam Performance

Introduction and Research Questions

In this chapter I discuss how students' study behavior is associated with student overall performance in Organic Chemistry 2. This chapter focuses on how each of the five main categories of student study behaviors: 1) In-class study behaviors, 2) Use of additional instructional resources, 3) Use of written/online resources, 4) Use of practice problems, and 5) use of group work were associated with overall performance. In particular a large emphasis is placed on comparing active versus passive uses of various resources.

This chapter will explore the following research questions:

1) How are students' study behaviors associated with their overall exam performance in a second semester organic chemistry course (Organic Chemistry 2)?
2) After identifying study behaviors significantly correlated with student performance, how does deviation from students' average usage of a resource impact their performance on exams?

Methods used in analysis, particularly the selection and creation of variables (though only chapter specific variables, many variables were previously discussed in Chapter 3) and the HLM models used in the analysis are described in the beginning of this chapter. Subject-mean centered covariates and time-dependent deviation variables were only developed for study behaviors, which had been found to be significantly correlated to students' overall performance in order avoid the creation of too many variables. This method of analysis also provided information on whether addition of a given study behavior was associated with exam performance. The chapter then describes the results and conclusions of this analysis.

Methods: Participants and Variables

There were 469 participants included in this study; the population consisted of any students who had completed at least one of the surveys. students who only completed the pilot survey were not included in the analysis. On average, participants completed 2.3 surveys. The majority of the variables used in this analysis were described in Chapter 3 in the Variable section. This section will describe the response variable used in this analysis and the subject-mean centered covariates developed to answer research question 2.

Response variable:

pExam: A continuous variable corresponding to student's percentage score on the exam. This variable was constructed by dividing students' raw exam scores by the total number of points available in the exam and then multiplying by 100. Percentage scores were used to normalize between exams since each of the three exams had different total numbers of points.

Subject-mean and mean-centered variables:

m_study_hours_10+hr: A continuous level-2 covariate corresponding to the fraction of occasions that students studied more than 10 hours prior to taking their exam. This variable was constructed by taking the mean of the level-1 dummy variable **studied_10hrs** which corresponded to whether or not an individual had studied more than 10 hours at a given time point. This provides a measurement of whether or not a student generally studied more than 10 hours.

dev_study_hours_10+hr: A continuous level-1 covariate corresponding to the difference between whether or not a student studied more than 10 hours at a given occasion and the fraction of times that a student studied more than 10 hours. This variable was constructed by subtracting m_studied_10hrs from studied_10hrs at a given occasion. This provides a measure for change in whether or not a student studied more than 10 hours at a given occasion.

m_notes_ownwords: A continuous level-2 covariate corresponding to the fraction of occasions that students engaged in active note-taking. This variable was constructed by taking the mean of the level-1 dummy variable **notes_ownwords** which corresponded to whether or not an individual had engaged in active note-taking at a given time point. This provides a measurement of whether or not a student generally engaged in active note-taking.

dev_notes_ownwords: A continuous level-1 covariate corresponding to the difference between whether or not a student engaged in active note-taking at a given occasion and the fraction of times that a student engaged in active note-taking. This variable was constructed by subtracting **m_notes_ownwords** from **notes_ownwords** at a given occasion. This provides a measure for change in whether or not a student engaged in active note-taking at a given occasion.

m_problem_testunderstanding: A continuous level-2 covariate corresponding to the fraction of occasions that students predominantly tested their understanding of the material when working on practice problems. This variable was constructed by taking the mean of the level-1 dichotomous variable **problem_testunderstanding**. This provides a measurement of whether or not a student generally tested their understanding while working on practice problems.

dev_problem_testunderstanding: A continuous level-1 covariate corresponding to the difference between whether or not a student predominantly tested their understanding at a given occasion and the fraction of times that a student predominantly tested their understanding. This variable was constructed by subtracting **m_problem_testunderstanding** from **problem_testunderstanding** at a given occasion. This provides a measure for change in whether or not a student engaged in active note-taking at a given occasion.

m_grpstudy_passive: A continuous level-2 covariate corresponding to the fraction of occasions that students engaged in passive group work. This variable was constructed by taking the mean of the level-1 dummy variable corresponding to passive group work. This provides a measurement of whether or not a student generally engaged in passive group work.

dev_grpstudy_passive: A continuous level-1 covariate corresponding to the difference between whether or not a student engaged in passive group work at a given occasion and the fraction of times that a student engaged in passive group work. This variable was constructed by subtracting **m_grpstudy_passive** from **grp_studypassive** at a given occasion. This provides a measure for change in whether or not a student engaged in passive groupwork at a given occasion or alternatively either worked alone or engaged in active group work.

Methods: Statistical Analysis

The longitudinal design of the study led to the use of Hierarchical Linear Modeling (HLM) in data analysis. HLM is a regression-based technique that can be used when data is clustered in some way (Rabe-Hesketh & Skrondal, 2012). The typical example is students who are clustered within schools. HLM can also be used with longitudinal data and treats occasions as nested within individuals. To answer the research questions two 2-level hierarchical models were used.

Model 1 was used to answer the first research question; which centered on how study behaviors were associated with overall exam scores. No Level-2 time-invariant variables were used in the first model; all explanatory variables were Level-1 time-varying variables.

Model 1: Study behaviors

- **Level-1 within-student effects**

  $pExam_{ij} = \beta_{0j} + \beta_1(occasion\_2)_i + \beta_2(occasion\_3)_i + \beta_3(study\_hours\_4\text{-}7hrs)_{ij} + \beta_4(study\_hours\_7\text{-}10hrs)_{ij} + \beta_5(study\_hours\_10{+}hr)_{ij} + \beta_6(lecture\_collaboration)_{ij} + \beta_7(worksheet\_revision)_{ij} + \beta_8(notes\_ownswords)_{ij} + \beta_9(review\_session)_{ij} + \beta_{10}(office\_hours)_{ij} + \beta_{11}(piazza\_passive)_{ij} + \beta_{12}(piazza\_active)_{ij} + \beta_{13}(review\_memorize)_{ij} + \beta_{14}(review\_Ptypes)_{ij} + \beta_{15}(review\_testunderstanding)_{ij} + \beta_{16}(review\_connectideas)_{ij} + \beta_{17}(review\_concepts)_{ij} + \beta_{18}(review\_Psolving)_{ij} + \beta_{19}(problem\_memorize)_{ij} + \beta_{20}(problem\_Ptypes)_{ij} + \beta_{21}(problem\_testunderstanding)_{ij} + \beta_{22}(problem\_connectideas)_{ij} + \beta_{23}(problem\_concepts)_{ij} + \beta_{24}(problem\_Psolving)_{ij} + \beta_{25}(grpstudy\_passive)_{ij} + \beta_{26}(grpstudy\_active)_{ij} + \varepsilon_{ij};\ \varepsilon_{ij} \sim N(0,\theta)$

- **Level 2 between student effects**

  $\beta_{0j} = \gamma_{00} + u_{0j};\ u_{0j} \sim N(0,\psi)$

Exam score at specific occasions (i) for a specific student (j) are modeled, using a random effects model with robust standard errors, as a function of a student-specific intercept (β_{0j}), indicators for occasion, indicators for student time spent studying (4-7 hours, 7-10 hours, 10+ hrs), and indicators of students study behaviors (in-class behaviors, use of additional instructional resources, use of written/online resources, use of practice problems, and use of group work). The u_{0j} in the intercept represents a normally distributed student-specific error term with mean 0 and variance ψ, and ε_{ij} represents a normally distributed student and occasion specific error term with mean 0 and variance θ, controlling for covariates.

After identifying study behaviors associated with overall exam scores I decided to further explore the results. To do this I developed a second model to answer research question 2. This model utilized subject means and deviations from subject means for study behaviors found to be significantly associated with overall exam scores. Use of subject means and deviations allowed me to distinguish between the effects of average usage and deviations in usage; this can also be described as splitting the variable into between-student and within-student components. Use of subject means and deviations from subject means can provide more information about the nature of the associations seen in Model 1 and some preliminary information about causality. A model consisting entirely of subject means and deviations from subject means would have required the testing of too many variables so a preliminary model was needed to focus the analysis to the most relevant study behavior (those significantly associated with exam scores).

Four significant study behaviors study_hours_10+hr, notes_ownwords, problem_testunderstanding, and grpstudy_passive were associated with exam scores and converted into subject mean and deviation variables. The remaining study behaviors were included in the model with no modifications. Level 2-time independent variables consisted of m_study_hours_10+hr, m_notes_ownwords, m_problem_testunderstanding, grpstudy_passive. Level 1 variables consisted of dev_study_hours_10+hr, dev_notes_ownwords, dev_problem_testunderstanding, dev_grpstudy_passive, all remaining non-significant study behaviors variables, and the occasion control variables.

Model 2: Between and within effects (Subject means and deviations from subject means are shown in bold)

- **Level-1 within-student effects**
  $pExam_{ij} = \beta_{0j} + \beta_1(\text{occasion\_2})_i + \beta_2(\text{occasion\_3})_i + \beta_3(\text{study\_hours\_4-7hrs})_{ij} + \beta_4(\text{study\_hours\_7-10hrs})_{ij} + \boldsymbol{\beta_5}(\textbf{dev\_study\_hours\_10+hr})_{ij} + \beta_6(\text{lecture\_collaboration})_{ij} + \beta_7(\text{worksheet\_revision})_{ij} + \boldsymbol{\beta_8}(\textbf{dev\_notes\_ownswords})_{ij} + \beta_9(\text{review\_session})_{ij} + \beta_{10}(\text{office\_hours})_{ij} + \beta_{11}(\text{piazza\_passive})_{ij} + \beta_{12}(\text{piazza\_active})_{ij} + \beta_{13}(\text{review\_memorize})_{ij} + \beta_{14}(\text{review\_Ptypes})_{ij} + \beta_{15}(\text{review\_testunderstanding})_{ij} + \beta_{16}(\text{review\_connectideas})_{ij} + \beta_{17}(\text{review\_concepts})_{ij} + \beta_{18}(\text{review\_Psolving})_{ij} + \beta_{19}(\text{problem\_memorize})_{ij} + \beta_{20}(\text{problem\_Ptypes})_{ij} + \boldsymbol{\beta_{21}}(\textbf{dev\_problem\_testunderstanding})_{ij} + \beta_{22}(\text{problem\_connectideas})_{ij} + \beta_{23}(\text{problem\_concepts})_{ij} + \beta_{24}(\text{problem\_Psolving})_{ij} + \boldsymbol{\beta_{25}}(\textbf{dev\_grpstudy\_passive})_{ij} + \beta_{26}(\text{grpstudy\_active})_{ij} + \varepsilon_{ij};\ \varepsilon_{ij} \sim N(0,\theta)$

- **Level 2 between student effects**

  $\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{m\_study\_hours\_10+hr}) + \gamma_{02}(\text{m\_notes\_ownwords}) + \gamma_{03}(\text{m\_problem\_testunderstanding}) + \gamma_{04}(\text{m\_grpstudy\_passive}) + u_{0j};\ u_{0j} \sim N(0,\psi)$

Exam score at specific occasions (i) for a specific student (j) were modeled, using a random effects model with robust standard errors, as a function of a student-specific intercept (β_{0j}), indicators for occasion, indicators for student time spent studying (4-7 hours, 7-10 hours, 10+ hrs), indicators of students study behaviors (in-class behaviors, use of additional instructional resources, use of written/online resources, use of practice problems, and use of group work, and deviation scores). Student-specific intercepts were modeled as a function of subject-level variables, i.e., the fraction of the time they used the behaviors study_hours_10+hr, notes_ownwords, and problem_testunderstanding, and grpstudy_passive. The u_{0j} in the intercept represents a normally distributed student-specific error term with mean 0 and variance ψ, and ε_{ij} represents a normally distributed student and occasion specific error term with mean 0 and variance θ, controlling for covariates.

A null model was also used to determine the intraclass correlation coefficient. Null models, or empty models, refer to models with no independent variables and can be used to determine variance at the student and occasion level. All HLM analyses were done using Stata 13 software.

Results: Descriptive Statistics

While 538 students were registered in Organic Chemistry 2, only 87% of students were included as participants in analysis (*N* = 469). This corresponded to a total of 1084 responses, with respondents completing an average of 2.3 surveys (*Min* = 1, *Max* = 3). 488 individuals completed at least one survey, however participants were dropped from the analysis if they had no reported exam scores as this indicated either a very early drop of the course or that they had mis-reported their student ID in the survey. Responses were then dropped if Level-1 variables included in the model were missing.

The mean normalized exam score, pExam, was 69.17. pExam appears to vary more widely between students than within students (between exams for the same student) (See Table 5.1). Figures 5.1 shows the histogram for pExam and indicates a left-skew. However, analysis of level-1 and level-2 residuals (See Appendix D) indicate normal distributions so no adjustments were made to the response variable.

Students on average studied more than 10 hours on 56% of the time, actively took notes 63% of the time, and predominantly tested their understanding when working on problems 57% of the time. Students however only engaged in passive group studying 8% of the time which indicates that this is a fairly rare behavior.

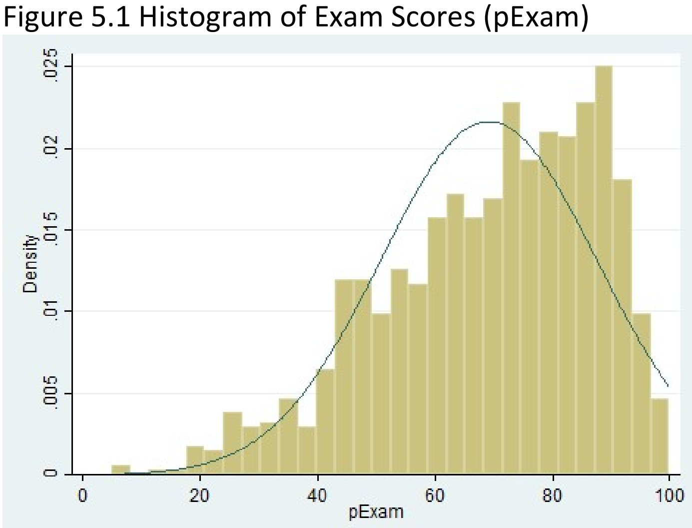
Table 5.1

Mean and standard deviation, within and between students, for continuous variables

Variable		Mean*	Standard Deviation
pExam	Overall	69.17	18.44
	Between		17.69
	Within		7.20
m_study_hours_10+hr	Overall	0.57	0.40
	Between		
	Within		
dev_study_hours_10+hr	Overall	0.00	0.29
	Between		
	Within		
m_notes_ownwords	Overall	0.63	0.40
	Between		
	Within		
dev_notes_ownwords	Overall	0.00	0.26
	Between		
	Within		
m_problem_testunderstanding	Overall	0.58	0.37
	Between		
	Within		
dev_problem_testunderstanding	Overall	0.00	0.33
	Between		
	Within		
m_grpstudy_passive	Overall	0.08	0.20
	Between		
	Within		
dev_grpstudy_passive	Overall	0.00	0.18
	Between		
	Within		

*Units for pExam were percentage points, for subjects means were fraction of occasions, and for deviations from subject means were difference between the usage at a given occasion and fraction of total occasions the study behavior was used.

Figure 5.1 Histogram of Exam Scores (pExam)



Results: Hierarchical and Linear Modeling

Two models and a null model were used to answer the research questions. This section will discuss the results from these three models. First outcomes from the null model, particularly the intraclass correlation coefficient (ICC) will be reported. The intraclass correlation coefficient is a descriptive statistic used to describe how strongly units in the same group resemble each other. A higher ICC indicates that use of HLM is appropriate. Then results for the two models will be discussed. This section is organized such that each sub-section focuses on different category of study behaviors. First the study behaviors significantly associated with exam scores in Model 1 are discussed and then the results from Model 2 are discussed to provide additional information on how mean usage of the resource and deviations from the mean are associated with exam scores. Results for the two models are also summarized in Tables 5.2 and 5.3.

The null models indicated that there was a large variance between students for normalized exam scores (See Table 5.2). The high intraclass correlation (ICC) for normalized exam scores, estimated as 0.74, suggests that 74% of the variance in exam scores is due to student level-characteristics.

Control variables

Results from Model 1 indicated significant differences in exam scores at the three different occasions. Students on average scored significantly higher on exams at occasion 2 and occasion 3 than at occasion 1 (See Table 5.2). Exam scores at occasion 2 were also on average estimated to be 2.85 points higher than exams scores at occasion 3 ($p<0.001$). This suggests that there were significant differences between the difficulty of the three exams relative to the students' abilities.

Study hours

Model 1 showed, after controlling for study behaviors, that students who studied more than 10 hours the week before the exam scored on average an estimated 3.16 points higher on the exam than those who studied <4 hours (See Table 5.2). Students who studied more than 10 hours also scored on average an estimated 3.36 points higher than those who studied 4-7 hours ($p<0.01$) and 2.36 points higher than those who studied 7-10 hours ($p<0.01$). Students who studied 7-10 hours did not perform significantly differently than those who studied 4-7 hours.

Model 2 was then used to further probe this association, by including the subject mean and deviation score for studying more than 10 hours (See Table 5.3). This analysis showed that students who studied more than 10 hours a larger proportion of the time were not estimated to perform significantly higher than those who studied more than 10 hours a smaller proportion of the time. However, students who change from not studying more than 10 hours to studying more than 10 hours, were estimated to on average perform 3.61 point higher ($p<0.05$).

Table 5.2

Fixed and Random Effects for the HLM Model 1

	Null Model pExam (100 total points)	Model 1 pExam (100 total points)
Fixed Part		
Intercept	67.88(0.81)***	55.08(2.70)***
Time (Reference: Occasion 1)		
Occasion 2		10.85(0.66)***
Occasion 3		7.99(0.63)***
Study Hours (Reference: <4 hours)		
4-7 hours		-0.19(1.53)
7-10 hours		0.81(1.56)
10 + hours		3.17(1.47)**
In-Class Study Behaviors (Reference: Did not engage in study behavior)		
Collaborated/Talked with neighbors during lecture		1.40(0.92)
Wrote down own thoughts/words when taking notes		2.03(0.81)**
Worked on problem, then revised the solution when the instructor went over the problem		0.84(0.88)
Use of additional Instructional Resources		
Review Session (Reference: Did not attend review session)		-0.51(0.82)
Office Hours (Reference: Did not attend office hours)		0.51(0.96)
Piazza (Reference: Did not use Piazza)		
Used Piazza passively		-0.82(0.98)
Used Piazza actively		0.89(1.19)
Use of written/online resources (Reference: Did not engage in study behavior)		
Memorized key words, concepts, molecules, and reactions		-0.69(0.88)
Reviewed problem types that might appear on quiz/exam		-0.81(0.87)
Tested understanding of material		0.84(0.86)
Built connections among ideas		0.72(0.99)
Identified underlying concepts		1.54(0.92)
Reviewed problem-solving approaches		-0.60(0.90)
Use of practice problems (Reference: Did not engage in study behavior)		
Memorized key words, concepts, molecules, and reactions		0.40(0.84)
Reviewed problem types that might appear on quiz/exam		1.03(0.83)
Tested understanding of material		1.88(0.85)**
Built connections among ideas		0.61(1.05)
Identified underlying concepts		0.97(0.84)
Reviewed problem-solving approaches		1.51(0.84)*
Group Study (Reference: Did not engage in group work)		
Passively engaged in groupwork		-3.53(1.42)**
Actively engaged in groupwork		-0.89(0.88)
Random Part		
Between student variance ψ	258.97	244.54
Within student variance θ	92.01	58.66

*-p-value approaching significance **p-value<0.05 ***p-value<0.001

Table 5.3

Fixed and Random Effects for the HLM Model 2

	Null Model pExam (100 total points)	Model 2 pExam (100 total points)
Fixed Part		
Intercept	67.88(0.81)***	53.72(3.32)***
Time (Reference: Occasion 1)		
Occasion 2		10.90(0.66)***
Occasion 3		7.92(0.63)***
Study Hours (Reference: <4 hours)		
4-7 hours		-0.18(1.53)
7-10 hours		0.85(1.57)
Average of 10+ hours		**1.29(2.17)**
Deviation in 10+ hours		**3.61(1.52)****
In-Class Study Behaviors (Reference: Did not engage in study behavior)		
Collaborated/Talked with neighbors during lecture		1.25(0.90)
Average of Wrote down own thoughts/words when taking notes		**4.00(1.93)****
Deviation in Wrote down own thoughts/words when taking notes		**1.44(0.88)**
Worked on problem, then revised the solution when the instructor went over the problem		0.65(0.88)
Use of additional Instructional Resources		
Review Session (Reference: Did not attend review session)		-0.42(0.82)
Office Hours (Reference: Did not attend office hours)		0.53(0.96)
Piazza (Reference: Did not use Piazza)		
Used Piazza passively		-0.72(0.98)
Used Piazza actively		0.91(1.18)
Use of written/online resources (Reference: Did not engage in study behavior)		
Memorized key words, concepts, molecules, and reactions		-0.71(0.88)
Reviewed problem types that might appear on quiz/exam		-0.76(0.87)
Tested understanding of material		0.71(0.86)
Built connections among ideas		0.61(0.98)
Identified underlying concepts		1.34(0.92)
Reviewed problem-solving approaches		-0.61(0.90)
Use of practice problems (Reference: Did not engage in study behavior)		
Memorized key words, concepts, molecules, and reactions		0.34(0.84)
Reviewed problem types that might appear on quiz/exam		0.95(0.82)
Average of Tested understanding of material		**6.02(2.12)*****
Deviation in Tested understanding of material		**-1.20(0.89)**
Built connections among ideas		0.50(1.03)
Identified underlying concepts		0.83(0.84)
Reviewed problem-solving approaches		1.41(0.83)
Group Study (Reference: Did not engage in group work)		
Average of Passively engaged in groupwork		**-13.47(3.49)*****
Deviation in Passively engaged in groupwork		**1.99(1.52)**
Actively engaged in groupwork		-0.77(0.87)
Random Part		
Between student variance ψ	258.97	235.50
Within student variance θ	92.01	58.26

*-p-value approaching significance **p-value<0.05 ***p-value<0.001

In-Class study behaviors

Model 1 indicated, after controlling for all other covariates, that only writing down one's own thoughts/words when taking notes was associated with higher exam performance (See Table 5.2). Students who did so were estimated to on average score 2.03 points higher than those who did not, controlling for all other covariates. Model 2 indicated that students who always wrote down their own thoughts/words when taking notes were estimated to on average perform 4.00 points higher than students who did not (See Table 5.3). However, there was no significant difference in students who changed in their usage of this strategy.

Use of additional instructional resources

Model 1 indicated that there were no instructional resources that were significantly associated with a higher overall exam score, controlling for all other covariates (See Table 5.2). Students who attended review sessions and office hours did not perform differently than students who did utilize these resources. Neither active nor passive use of piazza message boards was significantly associated with student exams scores. Students who actively used piazza were estimated to on average score 1.70 points higher on exams than students who passively used piazza but this association was only approaching significance ($p=0.074$)

Use of online/written resources

Students' exam scores were not significantly associated with their use of online or written resources, controlling for all other covariates (See Table 5.2).

Use of practice problems

Model 1 indicated that students who predominantly used practice problems to test their understanding of the materials were estimated to on average score 1.88 points higher on their exams than students who did not, controlling for all other covariates (See Table 5.2). Model 2 indicated that students who always tested their understanding when using practice problems were estimated to on average perform 6.02 points higher than students who never did so (See Table 5.3). However, there was no significant change for students who changed in their usage of this strategy.

Use of group work

Model 1 indicated that students who passively engaged in group work were estimated to on average score 3.53 points lower on exams than those who did not engage in group work controlling for all other covariates ($p<0.05$). Students who actively engaged in group work did not perform significantly differently than those who did not engage in groupwork at all, but did perform on average an estimated 2.71 points higher than students who passively engaged in group work though this result was only approaching significance (See Table 5.2).

Model 2 then indicated that students who on average engaged in passive group work were estimated to on average perform 13.47 points lower than students who did not engage in group work (See Table 5.3). However, there were no significant changes in mean score for students who changed from not engaging in groupwork to passively engaging in groupwork.

Data Analysis

This chapter sought to identify how student study behaviors were associated with their overall exam performance. It also sought to determine if change in usage of a strategy and/or overall consistency in usage of a strategy was associated with overall exam performance. Only four study behaviors were found to be associated with Organic Chemistry 2 exam scores: Spending more than 10 hours studying the week before the exam, active note-taking (writing down one's own thoughts/words), testing one's understanding while completing practice problems, and engaging in passive group work (See Table 5.2).

It was unsurprising that spending a high amount of time on task the week before an exam would be associated with high performance. What was more surprising however was that in the model with subject-mean centered covariates (Model 2) there was no significant difference between a student who tended to study more than 10 hours and one that tended not to study more than 10 hours before the exam. Significant differences arose however when a student changed the amount of time they spent studying for an exam. One possibility for this result is that students who choose to study consistently week to week may study fewer hours the week before the exam but still perform at a similar level. The finding that a student who changes from studying 10 hours or less to more than 10 hours the week before the exam, scores significantly higher, suggests that, for a given student, who may have constant habits in the intervening weeks, additional study just before an exam does help them to achieve higher scores.

While there has not been that much literature exploring the nuances of students' engagement with active-learning resources in the classroom, the finding that active note-taking in class would be associated with overall exam performance is consistent with principles of active learning. It was, however, somewhat more surprising that other in-class behaviors, such as collaborating with peers or working on problems, were not. It was also interesting that the fraction of times that a student actively took notes was significantly associated with their exam score but not changes in their usage of this study behavior. This suggests perhaps that it is important that students consistently write down their own thoughts/words when taking notes, and that simply doing so on one occasion was insufficient to impact exam scores. It may also suggest, however, that higher performing students are better able or more likely to engage in active note-taking. Future work will be needed to determine whether consistent usage of this strategy is helpful or simply representative of greater comfort with the material.

When asked about how they predominantly used practice problems, only students who predominantly tested their understanding performed significantly better. This would seem to preliminarily indicate that students that engage in metacognitive self-testing behavior while solving practice problems perform higher on Organic Chemistry 2 exams than those who do not. Further analysis indicates that there is a significant difference between students who on average tested their understanding while completing practice problems and those who did not, but there is again no difference between students who changed between testing their understanding and not testing their understanding. This can be potentially interpreted in two ways. One is that students who perform higher in the course are better able to or more likely to test their understanding while completing practice problems. The other is that consistent use of this study behavior is key and that simply using it in preparation for one exam is insufficient to have an impact.

One of the most surprising results was that students who engaged in passive group work (mostly listening) the week before the exam were estimated to perform significantly lower than students who did not engage in group work at all. Also interesting was that students who participated in active group work (asking and answering questions) did not perform significantly differently than those who did not engage in groupwork. While previous work such as that of Sinapuelas & Stacy (2015) has shown that the nature of students' group work impacts their performance it had not seemed to indicate that there were forms of groupwork that might be associated with lower gains than not engaging in group work. These results seem to indicate that the ways in which a student engages in group work are important and that some may less effective than forms of studying available to students outside of a group.

It was interesting that the results of Model 2 indicated that while the estimated average score was 13.47 points lower for a student who always engaged in passive group work than one who never engaged in group work there was no significant difference in scores when a student changed from one strategy to the other. This lack of significance may be, at least in part, due to the small sample size (see Appendix D). It may, however, also suggest that students who feel less comfortable with the material are more likely to engage in passive group work. It may also suggest that it is consistent usage of the strategy that is the problem. This makes sense if one considers that working with peers and simply listening is more akin to attending lecture than it is to working in a group. In this case by consistently attending group and listening a student is spending valuable study time on a less effective strategy.

It is also interesting to consider which study behaviors were not associated with overall exam performance. Neither usage of additional instructional resources or reviewing online/written resources were associated with exam performance. This may suggest that the variables as constructed did not capture the nuances of student usages of the strategies. This may be particularly true for additional instructional resources where the relatively low usage of office hours and review sessions meant that it was difficult to distinguish between the different types of usages. Also, observations of the review sessions seemed to indicate that they were normally conducted as lecture with little to no participation and tended to focus on lower level concepts. Overall these results may simply suggest that these study behaviors do not have as strong of an influence on overall student performance.

Conclusions

While this work does demonstrate significant associations between students' study behaviors and their performance in organic chemistry, there are still limitations to these results. The first is that the impact of many of these strategies is fairly small. This suggests that the instrument may not be capturing all relevant resources and study behaviors associated with them. The instrument is also limited by the fact that it's collecting student self-report data. Self-report data is limited by student memory and interpretations of questions. In order to mitigate recall bias students were asked to reflect on their studying the previous week during a time of focused exam preparation. The instrument was also piloted multiple times to try to ensure that items were being interpreted correctly (See Chapter 3).

The emphasis on students' study behaviors in the week leading up to exams however is another limitation. While asking students to reflect on the previous week and a time period of particular focused preparation reduces recall bias and can be interpreted as representative of main study efforts there are some limitations. Work such as Szu et al. (2011) has indicated that early and consistent studying is important to student performance and this is not captured by this instrument. Also, in Organic Chemistry 2 the material covered the week before the exam was not tested in the upcoming exam. This resulted in students often being less focused in-class the week prior to exams so in-class study behaviors may not be accurately captured. It's possible that students may participate more actively in class on other weeks and it may be helpful in future work to explore student study behaviors during non-exam preparatory times.

It's also difficult to determine causality, even when using subject-mean centered covariates. In cases where only the student mean is associated with overall performance it is difficult to know if the level of usage is important or if stronger students are more likely to utilize the strategy. Future work will be needed to distinguish whether these study behaviors lead to better outcomes.

Finally, another limitation is the assumption in this chapter that exams are mostly homogenous and that a given study behavior would help a student on the exam overall. This, as was discuss earlier in this dissertation, is almost certainly untrue as exams are made of a variety of different types of questions with different relationships to the curriculum. In order to truly understand how students' study behaviors influence student performance, it is important to explore how student study behaviors influence student performance on different components on the exam, specifically different categories of questions and questions with different relationships to the curriculum. These issues are explored in Chapters 6 and 7.

Even considering limitations however, the results of this chapter show that students who consistently engage in active note-taking and test their understanding of the material while working on practice problems performed better in Organic Chemistry 2 exams. It also showed that students who consistently engage in passive group work on average score lower than those who do not engage in group work. However, changes in usage of these behaviors was not associated with higher outcomes. Informing students of these associations at the beginning of the semester may be helpful but results seem to indicate that a change in behavior later in the semester may not have a significant impact. Only changes in whether one studied less than 4 hours or more than 10 hours seemed to be associated with overall exam performance. This suggests that if a student is struggling, increasing the amount of time studying may lead to improved exam scores. Use of early assessments on students' usage of the other significant study behavior may also help in determining which students may be more at risk in Organic Chemistry 2.

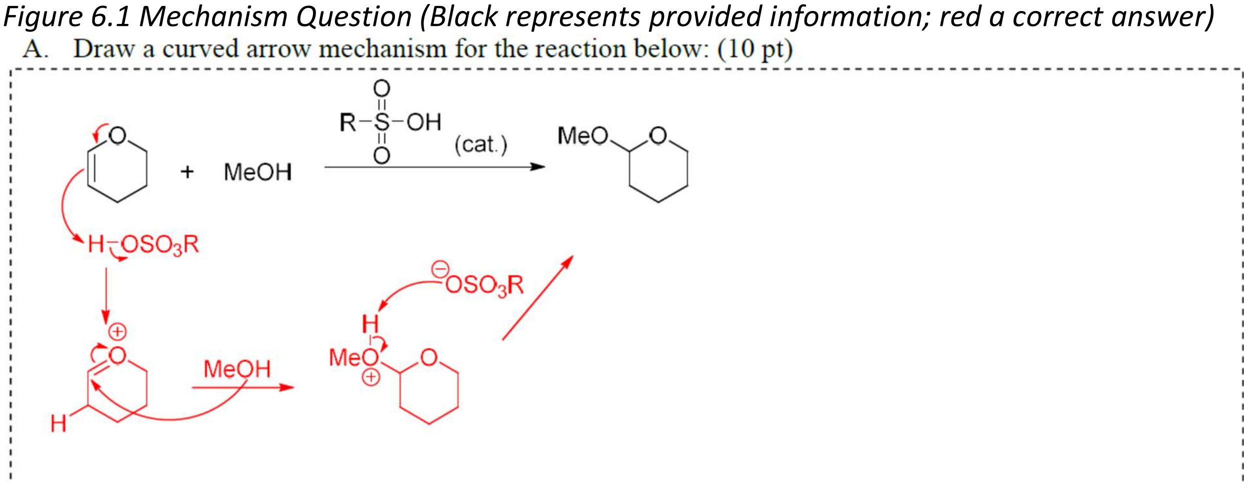
Chapter 6: Study Behaviors and Performance on Specific Questions Categories

Introduction and Research Questions

Chapter 5 identified four study behaviors associated with overall exam performance in Organic Chemistry 2. However as previously discussed, exams are made of a variety of different types of questions and it's important to determine if different question categories are associated with the same study behaviors as each other and with overall exam performance. While Organic Chemistry instructors will vary somewhat in what they teach and how they teach it, Organic Chemistry is notable for having various standardized categories of questions used in assessments (See Chapter 3 for more information on the types of problems used in Organic Chemistry 2) that are unique to the discipline. Two notable and ubiquitous questions used in organic chemistry assessments are mechanism questions and predict-the-product (PTP) questions.

Mechanism questions are a common area of study in the organic chemistry education literature (Bhattacharyya & Bodner, 2005). In a mechanism question, students are provided with starting materials and final products and asked to indicate movement of electrons and atoms that would lead to the product (See Figure 6.1). Students are responsible for demonstrating the movement of electrons via electron-pushing formalism (EPF) and all relevant intermediates. Electron-pushing formalism, as seen in Figure 6.1, is a form of notation in which movements of electron pairs are represented via curved arrows. Overall mechanisms test that students can explain reactivity that occurs. Specifically, that they understand what rearrangements are present when a given reaction occurs and that they are familiar with reaction steps.

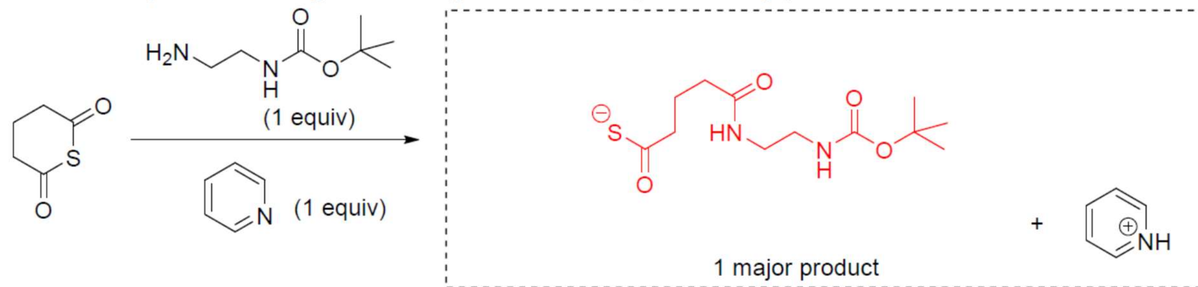
Figure 6.1 Mechanism Question (Black represents provided information; red a correct answer)



Predict the product (PTP) questions have been explored much less in the literature, though the related category of synthesis questions has been studied by Flynn et al. (2014). In contrast to mechanism questions, PTP questions provide students with starting materials and ask that they predict the final products (See Figure 6.2). Overall PTP questions test that students are familiar with patterns of reactivity and specific reactions and can use this information predictably. Instructors often use them to ensure that students are familiar with particular types of reactions, though as will be seen in Chapters 7 and 8, they can be designed to be both exercise-type and problem-type questions. While mechanism and PTP questions are testing different skills, these skills can be interrelated. In think-aloud interviews for example it was not uncommon for participants to draw out a partial mechanism to help in predicting a product.

Figure 6.2 PTP Question (Black represents provided information; red a correct answer)

Predict the organic product(s) from the following reactions. Where relevant, show all stereoisomers. Pay attention to any information given in the product boxes. (5 x 6 = 30 pt)



This chapter will focus on answering the following research questions:

1) How are students' study behaviors associated with their performance on predict-the-product questions?
   a. After identifying study behaviors significantly correlated with student performance on PTP questions, how does deviation from students' average usage of a resource impact their performance on PTP questions?
2) How are students' study behaviors associated with their performance on mechanism questions?
   a. After identifying study behaviors significantly correlated with student performance on mechanism questions, how does deviation from students' average usage of a resource impact their performance on mechanism questions?
3) Are predict the product and mechanism questions associated with different study behaviors? Are study behaviors associated with overall performance also associated with specific question categories?

This chapter will start by discussing the methods used in analysis, particularly the variables and the HLM models used in the analysis (for more information on the creation and selection of study behavior variables see Chapter 3). Subject-mean centered covariates and time-dependent deviation variables were only developed for study behaviors that had been found to be significantly correlated to students' performance on either PTP or mechanism questions in order to avoid testing too many variables in statistical models. This method of analysis also allowed for interpretation focused on the identified strategies. The chapter then describes the results and conclusions of this analysis and compare it to the results seen for overall exam performance.

Methods: Participants and Variables

There were 469 participants were included in this study. This population was made up of the same participants as in the previous chapter. The population consisted of any students who had completed at least one of the surveys. Students who only completed the pilot survey were not included in the analysis. On average participants completed 2.3 surveys. The majority of the variables used in this analysis were described in Chapter 3 in the Variable section. This section will describe the response variables used in this analysis and the subject-mean centered covariates developed to further probe the within and between effects for study behaviors significantly associated with performance

Response variables

PTP_Score: A continuous variable corresponding to students' percentage score on the predict-the-product questions of the exam. This variable was constructed by selecting three predict-the-product questions from each exam (See Appendix E for questions and average class scores). This number was chosen because the second exam only contained three predict-the-product questions and I wanted the score to be comparable across exams. The questions that received the lowest and highest average scores were selected and then the third question was chosen to be a high-scoring question covering a different reaction than the other two questions. The choice to focus on higher scoring responses was made due to the second midterm PTP questions having high average scores. This variable was constructed by dividing students' raw PTP scores by the total number of points available in the exam and then multiplying by 100. Percentage scores were used to normalize between exams and to more easily compare with mechanism scores and overall exam scores.

Mechanism_Score: A continuous variable corresponding to student's percentage score on the mechanism questions of the exam. This variable was constructed by selecting four mechanism questions from each exam (See Appendix E for questions and average class scores). This number was chosen because the first exam only contained four mechanism questions and I wanted the score to be comparable across exams. Questions with the highest average score were dropped from the second midterm and final, since the first midterm had comparatively lower scores. This variable was constructed by dividing students' raw mechanism scores by the total number of points available in the exam and then multiplying by 100. Percentage scores were used to normalize between exams and to more easily compare with PTP scores and overall exam scores.

Subject-Mean and Mean-centered Centered Variables

m_study_hours_7-10hr: A continuous level-2 covariate corresponding to the fraction of occasions that students studied 7-10 hours prior to taking their exam. This variable was constructed by taking the mean of the level-1 dummy variable **study_hours_7-10hr** which corresponded to whether or not an individual had studied 7-10 hours at a given time point. This provides a measurement of whether or not a student generally studied 7-10 hours.

dev_study_hours_7-10hr: A continuous level-1 covariate corresponding to the difference between whether or not a student studied 7-10 hours at a given occasion and the fraction of times that a student studied 7-10 hours. This variable was constructed by subtracting **m_study_hours_7-10hr** from **study_hours_7-10hr** at a given occasion. This provides a measure for change in whether or not a student studied 7-10 hours at a given occasion.

m_study_hours_10+hr: A continuous level-2 covariate corresponding to the fraction of occasions that students studied more than 10 hours prior to taking their exam. This variable was constructed by taking the mean of the level-1 dummy variable **study_hours_10+hr** which corresponded to whether or not an individual had studied more than 10 hours at a given time point. This provides a measurement of whether or not a student generally studied more than 10 hours.

dev_study_hours_10+hr: A continuous level-1 covariate corresponding to the difference between whether or not a student studied more than 10 hours at a given occasion and the fraction of times that a student studied more than 10 hours. This variable was constructed by subtracting **m_study_hours_10+hr** from **study_hours_10hr** at a given occasion This provides a measure for change in whether or not a student studied more than 10 hours at a given occasion.

m_lecture_collaboration: A continuous level-2 covariate corresponding to the fraction of occasions that students collaborated during lecture. This variable was constructed by taking the mean of the level-1 dummy variable **lecture_collaboration.** This provides a measurement of whether or not a student generally talked/collaborated with neighbors during lecture.

dev_lecture_collaboration: A continuous level-1 covariate corresponding to the difference between whether or not a student talked/collaborated with neighbors at a given occasion and the fraction of times that a student talked/collaborated. This variable was constructed by subtracting **m_lecture_collaboration** from **lecture_collaboration** at a given occasion. This provides a measure for change in whether or not a student talked/collaborated with neighbors at a given occasion.

m_worksheet_revision: A continuous level-2 covariate corresponding to the fraction of occasions that students worked on problems and then revised the solution. This variable was constructed by taking the mean of the level-1 dummy variable **worksheet_revision.** This provides a measurement of whether or not a student generally worked on problems and then revised during lecture.

dev_worksheet_revision: A continuous level-1 covariate corresponding to the difference between whether or not a student worked on problems and then revised at a given occasion and the fraction of times that a student worked on problems and then revised during lecture. This variable was constructed by subtracting **m_worksheet_revision** from **worksheet_revision**

at a given occasion. This provides a measure for change in whether or not a student talked/collaborated with neighbors at a given occasion.

m_review_connectideas: A continuous level-2 covariate corresponding to the fraction of occasions that students predominantly built connections among ideas when reviewing online or written resources. This variable was constructed by taking the mean of the level-1 dichotomous variable **review_connectideas**. This provides a measurement of whether or not a student generally built connections among ideas while reviewing online or written resources.

dev_review_connectideas: A continuous level-1 covariate corresponding to the difference between whether or not a student predominantly built connections among ideas at a given occasion and the fraction of times that a student predominantly built connections among ideas. This variable was constructed by subtracting **m_review_connectideas** from **review_connectideas** at a given occasion. This provides a measure for change in whether or not a student predominantly built connections among ideas while reviewing online or written resources.

m_problem_testunderstanding: A continuous level-2 covariate corresponding to the fraction of occasions that students predominantly tested their understanding of the material when working on practice problems. This variable was constructed by taking the mean of the level-1 dichotomous variable **problem_testunderstanding**. This provides a measurement of whether or not a student generally tested their understanding while working on practice problems.

dev_problem_testunderstanding: A continuous level-1 covariate corresponding to the difference between whether or not a student predominantly tested their understanding at a given occasion and the fraction of times that a student predominantly tested their understanding. This variable was constructed by subtracting **m_problem_testunderstanding** from **problem_testunderstanding** at a given occasion. This provides a measure for change in whether or not a student predominantly tested their understanding while working on practice problems.

m_problem_connectideas: A continuous level-2 covariate corresponding to the fraction of occasions that students predominantly built connections among ideas when working on practice problems. This variable was constructed by taking the mean of the level-1 dichotomous variable **problem_connectideas**. This provides a measurement of whether or not a student generally built connections among ideas while working on practice problems.

dev_problem_connectideas: A continuous level-1 covariate corresponding to the difference between whether or not a student predominantly built connections among ideas at a given occasion and the fraction of times that a student predominantly built connections among ideas. This variable was constructed by subtracting **m_problem_connectideas** from **problem_connectideas** at a given occasion. This provides a measure for change in whether or not a student predominantly built connections among ideas while working on practice problems.

m_problem_concepts: A continuous level-2 covariate corresponding to the fraction of occasions that students predominantly identified underlying concepts when working on practice problems. This variable was constructed by taking the mean of the level-1 dichotomous variable **problem_concepts**. This provides a measurement of whether or not a student generally identified underlying concepts while working on practice problems.

dev_problem_concepts: A continuous level-1 covariate corresponding to the difference between whether or not a student predominantly identified underlying concepts at a given occasion and the fraction of times that a student identified underlying concepts. This variable was constructed by subtracting **m_problem_concepts** from **problem_concepts** at a given occasion. This provides a measure for change in whether or not a student predominantly identified underlying concepts while working on practice problems.

m_problem_Psolving: A continuous level-2 covariate corresponding to the fraction of occasions that students predominantly reviewed problem-solving approaches when working on practice problems. This variable was constructed by taking the mean of the level-1 dichotomous variable **problem_Psolving**. This provides a measurement of whether or not a student generally reviewed problem-solving approaches while working on practice problems.

dev_problem_Psolving: A continuous level-1 covariate corresponding to the difference between whether or not a student predominantly reviewed problem-solving approaches at a given occasion and the fraction of times that a student reviewed problem-solving approaches. This variable was constructed by subtracting **m_problem_Psolving** from **problem_Psolving** at a given occasion. This provides a measure for change in whether or not a student reviewed problem-solving approaches while working on practice problems.

m_grpstudy_passive: A continuous level-2 covariate corresponding to the fraction of occasions that students engaged in passive group work. This variable was constructed by taking the mean of the level-1 dummy variable corresponding to passive group work **grpstudy_passive**. This provides a measurement of whether or not a student generally engaged in passive group work.

dev_grpstudy_passive: A continuous level-1 covariate corresponding to the difference between whether or not a student engaged in passive group work at a given occasion and the fraction of times that a student engaged in passive group work. This variable was constructed by subtracting **m_grpstudy_passive** from **grp_study_passive** at a given occasion. This provides a measure for change in whether or not a student engaged in passive groupwork at a given occasion or alternatively either worked alone or engaged in active group work.

Methods: Statistical Analysis

To answer the research questions four 2-level hierarchical models were used. Models 1 and 2 were used to answer the first research question; which centered on how various study behaviors were associated with PTP scores. No Level-2 time-invariant variables were used in the first model; all explanatory variables were Level-1 time-varying variables.

Model 1: Study behaviors and Predict-the-Product Score

- **Level-1 within-student effects**
  $PTP\_Score_{ij} = \beta_{0j} + \beta_1(occasion\_2)_i + \beta_2(occasion\_3)_i + \beta_3(study\_hours\_4\text{-}7hrs)_{ij} + \beta_4(study\_hours\_7\text{-}10hrs)_{ij} + \beta_5(study\_hours\_10\text{+}hr)_{ij} + \beta_6(lecture\_collaboration)_{ij} + \beta_7(worksheet\_revision)_{ij} + \beta_8(notes\_ownswords)_{ij} + \beta_9(review\_session)_{ij} + \beta_{10}(office\_hours)_{ij} + \beta_{11}(piazza\_passive)_{ij} + \beta_{12}(piazza\_active)_{ij} + \beta_{13}(review\_memorize)_{ij} + \beta_{14}(review\_Ptypes)_{ij} + \beta_{15}(review\_testunderstanding)_{ij} + \beta_{16}(review\_connectideas)_{ij} + \beta_{17}(review\_concepts)_{ij} + \beta_{18}(review\_Psolving)_{ij} + \beta_{19}(problem\_memorize)_{ij} + \beta_{20}(problem\_Ptypes)_{ij} + \beta_{21}(problem\_testunderstanding)_{ij} + \beta_{22}(problem\_connectideas)_{ij} + \beta_{23}(problem\_concepts)_{ij} + \beta_{24}(problem\_Psolving)_{ij} + \beta_{25}(grpstudy\_passive)_{ij} + \beta_{26}(grpstudy\_active)_{ij} + \varepsilon_{ij}; \quad \varepsilon_{ij} \sim N(0,\theta)$

- **Level 2 between student effects**

  $\beta_{0j} = \gamma_{00} + u_{0j}; \quad u_{0j} \sim N(0,\psi)$

PTP Scores at specific occasions (i) for a specific student (j) are modeled, using a random effects model with robust standard errors, as a function of a student-specific intercept (β_{0j}), indicators for occasion, indicators student time spent studying (4-7 hours, 7-10 hours, 10+ hrs), and indicators of students study behaviors (in-class behaviors, use of additional instructional resources, use of written/online resources, use of practice problems, and use of group work). The u_{0j} in the intercept represents a normally distributed student-specific error term with mean 0 and variance ψ, and ε_{ij} represents a normally distributed student and occasion specific error term with mean 0 and variance θ, controlling for covariates.

After identifying study behaviors associated with PTP scores a second model was developed to further answer research question 1. This model utilized subject means and subject--mean centered covariates for study behaviors found to be significantly associated with exam scores in the previous model. Use of subject means and deviations allowed me to distinguish between the effects of average usage and deviations in usage; this can also be described as splitting the variable into between-student and within-student components.

A model consisting entirely of subject means and deviations from subject means would have required the testing of too many variables so a preliminary model was needed to focus the analysis to the most relevant study behavior (those significantly associated with exam scores). Seven significant study behaviors: study_hours_7-10hr, study_hours_10+hr, worksheet_revision, problem_testunderstanding, problem_connectideas, problem_concepts, and grpstudy_passive were associated with exam scores and converted into subject mean and deviation variables. The remaining study behaviors were included in the model with no modifications.

Level 2 time-invariant variables consisted of m_study_hours_7-10hr, m_study_hours_10+hr, m_worksheet_revision, , m_problem_testunderstanding, m_problem_connectideas, m_problem_concepts, and m_grpstudy_passive. Level 1 variables consisted of dev_lecture_collaboration, dev_problem_connectideas, dev_problem_Psolving, and dev_grpstudy_passive, all remaining non-significant study behaviors variables, and the occasion control variables.

Model 2: Predict-the-Product Score Between and within effects (Subject means and deviations from subject means are shown in bold)

- **Level-1 within-student effects**

  PTP_Score$_{ij}$ = β_{0j} + β_1(occasion_2)$_i$ + β_2(occasion_3)$_i$ + β_3(study_hours_4-7hrs)$_{ij}$ + **β_4(dev_study_hours_7-10hr)$_{ij}$** + **β_5(dev_study_hours_10+hr)$_{ij}$** + β_6(lecture_collaboration)$_{ij}$ + **β_7(dev_worksheet_revision)$_{ij}$** + β_8(notes_ownswords)$_{ij}$ + β_9(review_session)$_{ij}$ + β_{10}(office_hours)$_{ij}$ + β_{11}(piazza_passive)$_{ij}$ + β_{12}(piazza_active)$_{ij}$ + β_{13}(review_memorize)$_{ij}$ + β_{14}(review_Ptypes)$_{ij}$ + β_{15}(review_testunderstanding)$_{ij}$ + β_{16}(review_connectideas)$_{ij}$ + β_{17}(review_concepts)$_{ij}$ + β_{18}(review_Psolving)$_{ij}$ + β_{19}(problem_memorize)$_{ij}$ + β_{20}(problem_Ptypes)$_{ij}$ + **β_{21}(dev_problem_testunderstanding)$_{ij}$** + **β_{22}(dev_problem_connectideas)$_{ij}$ +** **β_{23}(dev_problem_concepts)$_{ij}$** + β_{24}(problem_Psolving)$_{ij}$ + **β_{25}(dev_grpstudy_passive)$_{ij}$** + β_{26}(grpstudy_active)$_{ij}$ + ε_{ij}; $\varepsilon_{ij} \sim N(0,\theta)$

- **Level 2 between student effects**

  β_{0j} = γ_{00} + γ_{01}(m_study_hours_7-10hr) + γ_{02}(m_study_hours_10+hr) + γ_{03}(m_worksheet_revision) + γ_{04}(m_problem_testunderstanding) + γ_{05}(m_problem_connectideas) + γ_{06}(m_problem_concepts) + γ_{07}(m_grpstudy_passive) + u_{0j}; $u_{0j} \sim N(0,\psi)$

PTP score at specific occasions (i) for a specific student (j) were modeled, using a random effects model with robust standard errors, as a function of a student-specific intercept (β_{0j}), indicators for occasion, indicators student time spent studying (4-7 hours, 7-10 hours, 10+ hrs), indicators of students' study behaviors (in-class behaviors, use of additional instructional resources, use of written/online resources, use of practice problems, and use of group work), and deviation scores. Student-specific intercepts were modeled as a function of subject-mean covariates for study_hours_7-10hr, study_hours_10+hr, worksheet_revision, problem_testunderstanding, problem_connectideas, problem_concepts, and grpstudy_passive. The u_{0j} in the intercept represents a normally distributed student-specific error term with mean 0 and variance ψ, and ε_{ij} represents a normally distributed student and occasion specific error term with mean 0 and variance θ, controlling for covariates.

Models 3 and 4 were used to answer the second research question; which centered on how various study behaviors were associated with Mechanism scores. No Level-2 time-independent variables were used in the Model 3; all explanatory variables were Level-1 time-dependent variables.

Model 3: Study behaviors and Mechanism Score

- **Level-1 within-student effects**

$$Mechanism\_Score_{ij} = \beta_{0j} + \beta_1(occasion\_2)_i + \beta_2(occasion\_3)_i + \beta_3(study\_hours\_4\text{-}7hrs)_{ij} + \beta_4(study\_hours\_7\text{-}10hrs)_{ij} + \beta_5(study\_hours\_10\text{+}hr)_{ij} + \beta_6(lecture\_collaboration)_{ij} + \beta_7(worksheet\_revision)_{ij} + \beta_8(notes\_ownswords)_{ij} + \beta_9(review\_session)_{ij} + \beta_{10}(office\_hours)_{ij} + \beta_{11}(piazza\_passive)_{ij} + \beta_{12}(piazza\_active)_{ij} + \beta_{13}(review\_memorize)_{ij} + \beta_{14}(review\_Ptypes)_{ij} + \beta_{15}(review\_testunderstanding)_{ij} + \beta_{16}(review\_connectideas)_{ij} + \beta_{17}(review\_concepts)_{ij} + \beta_{18}(review\_Psolving)_{ij} + \beta_{19}(problem\_memorize)_{ij} + \beta_{20}(problem\_Ptypes)_{ij} + \beta_{21}(problem\_testunderstanding)_{ij} + \beta_{22}(problem\_connectideas)_{ij} + \beta_{23}(problem\_concepts)_{ij} + \beta_{24}(problem\_Psolving)_{ij} + \beta_{25}(grpstudy\_passive)_{ij} + \beta_{26}(grpstudy\_active)_{ij} + \varepsilon_{ij}; \quad \varepsilon_{ij} \sim N(0,\theta)$$

- **Level 2 between student effects**

$$\beta_{0j} = \gamma_{00} + u_{0j}; \quad u_{0j} \sim N(0,\psi)$$

Mechanism Scores at specific occasions (i) for a specific student (j) are modeled, using a random effects model with robust standard errors, as a function of a student-specific intercept (β_{0j}), indicators for occasion, indicators student time spent studying (4-7 hours, 7-10 hours, 10+ hrs), and indicators of students study behaviors (in-class behaviors, use of additional instructional resources, use of written/online resources, use of practice problems, and use of group work). The u_{0j} in the intercept represents a normally distributed student-specific error term with mean 0 and variance ψ, and ε_{ij} represents a normally distributed student and occasion specific error term with mean 0 and variance θ, controlling for covariates.

After identifying study behaviors associated with Mechanism scores Model 4 was developed to further answer research question 2. This model utilized subject means of covariates and mean centered covariates for study behaviors found to be significantly associated with exam scores in the previous model. Four significant study behaviors: lecture_collaboration, review_testunderstanding, problem_Psolving, and grpstudy_passive were associated with exam scores and converted into subject mean and deviation variables. The remaining study behaviors were included in the model with no modifications. Level 2-time invariant variables consisted of m_lecture_collaboration, m_review_testunderstanding, m_problem_Psolving, and m_grpstudy_passive. Level 1 variables consisted of dev_lecture_collaboration, dev_review_testunderstanding, dev_problem_Psolving, dev_grpstudy_passive, all remaining non-significant study behaviors variables, and the occasion control variables.

Model 4: Mechanism Score Between and within effects (Subject means and deviations from subject means are **shown** in bold)

- **Level-1 within-student effects**

  $PTP\_Score_{ij} = \beta_{0j} + \beta_1(occasion\_2)_i + \beta_2(occasion\_3)_i + \beta_3(study\_hours\_4\text{-}7hrs)_{ij} + \beta_4(study\_hours\_7\text{-}10hr)_{ij} + \beta_5(study\_hours\_10{+}hr)_{ij} + \boldsymbol{\beta_6(dev\_lecture\_collaboration)_{ij}} + \beta_7(worksheet\_revision)_{ij} + \beta_8(notes\_ownswords)_{ij} + \beta_9(review\_session)_{ij} + \beta_{10}(office\_hours)_{ij} + \beta_{11}(piazza\_passive)_{ij} + \beta_{12}(piazza\_active)_{ij} + \beta_{13}(review\_memorize)_{ij} + \beta_{14}(review\_Ptypes)_{ij} + \beta_{15}(review\_testunderstanding)_{ij} + \boldsymbol{\beta_{16}(dev\_review\_connectideas)_{ij}} + \beta_{17}(review\_concepts)_{ij} + \beta_{18}(review\_Psolving)_{ij} + \beta_{19}(problem\_memorize)_{ij} + \beta_{20}(problem\_Ptypes)_{ij} + \beta_{21}(problem\_testunderstanding)_{ij} + \beta_{22}(problem\_connectideas)_{ij} + \beta_{23}(problem\_concepts)_{ij} + \boldsymbol{\beta_{24}(dev\_problem\_Psolving)_{ij}} + \boldsymbol{\beta_{25}(dev\_grpstudy\_passive)_{ij}} + \beta_{26}(grpstudy\_active)_{ij} + \varepsilon_{ij}; \quad \varepsilon_{ij} \sim N(0,\theta)$

- **Level 2 between student effects**

  $\beta_{0j} = \gamma_{00} + \gamma_{01}(m\_lecture\_collaboration) + \gamma_{02}(m\_review\_connectideas) + \gamma_{03}(m\_problem\_Psolving) + \gamma_{04}(m\_grpstudy\_passive) + u_{0j}; \quad u_{0j} \sim N(0,\psi)$

Mechanism score at specific occasions (i) for a specific student (j) were modeled, using a random effects model with robust standard errors, as a function of a student-specific intercept (β_{0j}), indicators for occasion, indicators student time spent studying (4-7 hours, 7-10 hours, 10+ hrs), indicators of students study behaviors (in-class behaviors, use of additional instructional resources, use of written/online resources, use of practice problems, and use of group work), and deviation scores. Student-specific intercepts were modeled as a function of subject means- of the indicators for lecture_collaboration, review_connectideas, problem_Psolving and grpstudy_passive. The u_{0j} in the intercept represents a normally distributed student-specific error term with mean 0 and variance ψ, and ε_{ij} represents a normally distributed student and occasion specific error term with mean 0 and variance θ, controlling for covariates.

Two null models were also used to determine the intraclass correlation coefficient. Null models, or empty models, refer to models with no independent variables and can be used to determine variance at the student and occasion level. All HLM analyses were done using Stata 13 software.

Results: Descriptive Statistics

There were 469 participants included in this analysis; this corresponded to individuals who had completed at least one survey, had at least one reported exam score, and that were not missing any level-1 variables. This corresponded to a total of 1084 responses, with respondents completing an average of 2.3 surveys (*Min* = 1, *Max* = 3).

Table 6.1

Mean and standard deviation of within and between levels for continuous response variables

Variable		Mean*	Standard Deviation
PTP_Score	Overall	74.40	26.58
	Between		23.72
	Within		14.85
Mechanism_Score	Overall	73.50	22.57
	Between		20.99
	Within		10.93

*Units for PTP_Score and Mechanism_Score were percentage points,

The mean normalized predict-the-product sub-score (PTP_Score) was 74.40. PTP_Score varies more widely between students than within students (See Table 6.1). This means that there is more variation on between students average PTP_Scores than there between a single students PTP scores across the three exams. The mean normalized mechanism score (Mechanism_Score) was 73.50 and varies more widely between students than within students. Figure 6.1 shows the histograms for PTP_Score and Mechanism_Score and indicates that both have a left-skew. However, analysis of level-1 and level-2 residuals (See Appendix D) indicate normal distributions so no adjustments were made to the response variables.

Figure 6.1 Histograms of PTP_Score and Mechanism_Score

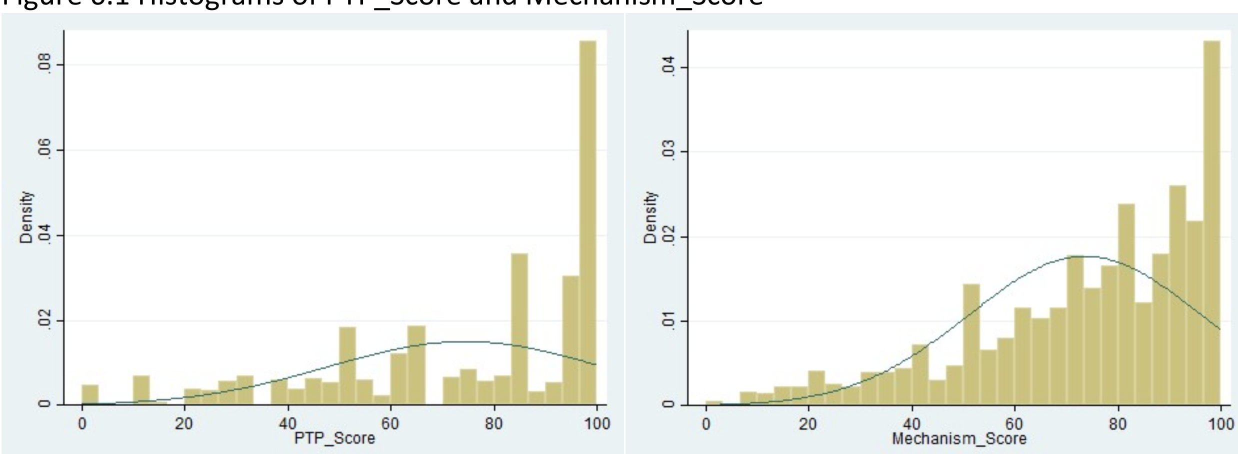


Table 6.2 lists the mean and standard deviation for the subject-mean centered variables developed for models 2 and 4. These results indicate that there is a range of usage for the different study behaviors. Students most frequent study behavior was working on lecture worksheets and then revising which they did on 72% of all occasion. They engaged in passive group-work least frequently and only did so on 8% of occasion. Standard deviations also tended to be comparable for mean and deviation variables, with mean variables usually having a slightly larger standard deviation.

Table 6.2

Mean and standard deviation of within and between levels for continuous variables

Variable		Mean*	Standard Deviation
m_study_hours_7-10hr	Overall	0.21	0.30
	Between		
	Within		
dev_study_hours_7-10hr	Overall	0.00	0.28
	Between		
	Within		
m_study_hours_10+hr	Overall	0.57	0.40
	Between		
	Within		
dev_study_hours_10+hr	Overall	0.00	0.29
	Between		
	Within		
m_lecture_collaboration	Overall	0.68	0.39
	Between		
	Within		
dev_lecture_collaboration	Overall	0.00	0.26
	Between		
	Within		
m_worksheet_revision	Overall	0.73	0.36
	Between		
	Within		
dev_worksheet_revision	Overall	0.00	0.26
	Between		
	Within		
m_review_connectideas	Overall	0.24	0.24
	Between		
	Within		
dev_review_connectideas	Overall	0.00	0.29
	Between		
	Within		
m_problem_testunderstanding	Overall	0.58	0.37
	Between		
	Within		
dev_problem_testunderstanding	Overall	0.00	0.33
	Between		
	Within		
m_problem_connectideas	Overall	0.12	0.23
	Between		
	Within		
dev_problem_connectideas	Overall	0.00	0.23
	Between		
	Within		
m_problem_concepts	Overall	0.34	0.36
	Between		
	Within		
dev_problem_concepts	Overall	0.00	0.31
	Between		
	Within		
m_problem_Psolving	Overall	0.39	0.38
	Between		
	Within		
dev_problem_Psolving	Overall	0.00	0.31
	Between		
	Within		
m_grpstudy_passive	Overall	0.08	0.21
	Between		
	Within		
dev_grpstudy_passive	Overall	0.00	0.18
	Between		
	Within		

*Units for subjects means were fraction of occasions, and for deviations from subject means were difference between the usage at a given occasion and fraction of total occasions the study behavior was used.

Results: Hierarchical and Linear Modeling

Four models and a two null model were used to answer the research questions. This section will discuss the results from these six models. First outcomes from the null models, particularly intraclass correlation coefficients, will be reported. Then results for the two models will be discussed. This section is organized such that each sub-section focuses on different category of study behaviors. First the study behaviors significantly associated with PTP Scores and Mechanism Scores are reported and then the results the subsequent follow-up models (Models 2 and 4) are discussed to provide additional information on how mean usage of the resource and deviations from the mean are associated with exam scores. Results from the four models are also summarized in Tables 6.3 and 6.4.

The null models indicated that there was a large between-student variance for PTP and mechanism scores (See table 6.3). There was also a high within-student variance for PTP scores; this variance was smaller for mechanism scores. There was a moderate intraclass correlation coefficient (ICC) for PTP scores, 0.45, which can be interpreted as that 45% of the variance in PTP scores is attributable to student level-characteristics. There was a moderate yet higher ICC for mechanism scores of 0.59, which can be interpreted as that 59% of the variance in mechanism scores is attributable to student-level characteristics.

Control variables

Results from Models 1 and 2 indicated significant differences in both PTP and Mechanism scores at the three occasions. Students on average scored significantly higher at occasion 2 and occasion 3 than at occasion 1 (See Table 6.3). Scores at occasion 2 were also on average estimated to be significantly higher than scores at occasion 3 for both Mechanism and PTP scores. PTP scores at occasion 2 were estimated to on average be 8.50 points higher than scores at occasion 3 ($p<0.001$). Mechanism scores were estimated to on average be 8.83 points higher. This suggests that there were significant differences between the difficulty of the three exams for the two question categories relative to students' abilities.

Study hours

Model 1 showed, after controlling for study behaviors, that students who studied 7-10 hours the week before the exam scored on average an estimated 5.74 points higher on predict-the-product questions than those who studied <4 hours (See Table 6.3). Students who studied more than 10 hours scored on average an estimated 5.48 points higher on predict the product questions than those who studied <4 hours. There were not significant differences between students who studied more than 10 hours, students who studied 7-10 hours, and students who studied 4-7 hours. In contrast, Model 3 indicated that, are controlling for all other covariates, students' hours spent studying were not significantly associated with scores on mechanism questions.

Table 6.3

Fixed and Random Effects for the HLM Models 1 & 3

	Null Model 1 PTP Score (100 total points)	Model 1 PTP Score (100 total points)	Null Model 3 Mechanism Score (100 total points)	Model 3 Mechanism Score (100 total points)
Fixed Part				
Intercept	73.39(1.04)***	52.38(5.05)***	72.31(0.94)***	57.36(4.23)***
Time (Reference: Occasion 1)				
Occasion 2		18.52(1.31)***		16.17(0.92)***
Occasion 3		10.02(1.43)***		7.34(1.05)***
Study Hours (Reference: <4 hours)				
4-7 hours		2.71(3.05)		0.27(2.18)
7-10 hours		5.48(2.69)**		0.61(2.25)
10 + hours		5.74(2.52)**		3.08(2.05)
In-Class Study Behaviors				
Collaborated/Talked with neighbors during lecture		1.09(1.75)		4.71(1.28)***
Wrote down own thoughts/words when taking notes		1.97(1.61)		1.64(1.18)
Worked on problem, then revised the solution when the instructor went over the problem		3.49(1.75)**		1.29(1.30)
Use of additional Instructional Resources				
Review Session		-2.76(1.67)*		-0.74(1.30)
Office Hours		3.31(1.82)*		0.12(1.26)
Piazza (Reference: Did not use Piazza)				
Used Piazza passively		-0.66(1.79)		-2.41(1.53)
Used Piazza actively		3.21(2.24)		-2.58(1.85)
Use of written/online resources				
Memorized key words, concepts, molecules, and reactions		-3.29(1.93)*		-0.65(1.46)
Reviewed problem types that might appear on quiz/exam		-2.66(1.81)		-0.70(1.35)
Tested understanding of material		-0.87(1.80)		1.36(1.33)
Built connections among ideas		-0.31(1.92)		3.39(1.49)**
Identified underlying concepts		-0.02(1.90)		1.64(1.46)
Reviewed problem-solving approaches		-2.10 (1.89)		0.32(1.44)
Use of practice problems				
Memorized key words, concepts, molecules, and reactions		2.83(1.72)		0.13(1.31)
Reviewed problem types that might appear on quiz/exam		2.99(1.69)*		0.24(1.22)
Tested understanding of material		6.08(1.57)***		1.95(1.21)
Built connections among ideas		5.00(2.00)**		-0.66(1.65)
Identified underlying concepts		3.39(1.69)**		0.45(1.17)
Reviewed problem-solving approaches		2.28(1.59)		3.00(1.25)**
Group Study (Reference: Did not engage in group work)				
Passively engaged in groupwork		-6.71(2.84)**		-7.35(2.19)**
Actively engaged in groupwork		-1.99(1.75)		-1.94(1.40)
Random Part				
Between student variance ψ	319.48	304.50	311.26	297.88
Within student variance θ	398.29	297.51	213.60	143.30

*-p-value approaching significance **p-value<0.05 ***p-value<0.001

Table 6.4

Fixed and Random Effects for the HLM Models 2 & 4 (Bold corresponds to new variables)

	Model 2 PTP Score (100 total points)	Model 4 Mechanism Score (100 total points)
Fixed Part		
Intercept	47.53(5.73)***	56.24(4.64)***
Time (Reference: Occasion 1)		
Occasion 2	18.74(1.32)***	16.14(0.91)***
Occasion 3	9.76(1.42)***	7.17(1.05)***
Study Hours (Reference: <4 hours)		
4-7 hours	3.27(3.02)	0.19(2.17)
7-10 hours		0.48(2.26)
Average of 7-10 hours	**6.42(4.41)**	
Deviation in 7-10 hours	**5.73(2.92)***	
10 + hours		3.03(2.05)
Average of 10+ hours	**4.66(3.70)**	
Deviation in 10+ hours	**6.92(2.73)****	
In-Class Study Behaviors		
Collaborated/Talked with neighbors during lecture	0.58(1.76)	
Average of Collaborated/Talked with neighbors during lecture		**6.83(2.58)*****
Deviation in Collaborated/Talked with neighbors during lecture		**3.96(1.46)*****
Wrote down own thoughts/words when taking notes	1.41(1.60)	1.40(1.19)
Worked on problem, then revised the solution when the instructor went over the problem		1.06(1.28)
Average of Worked on problem, then revised solution	**8.80(2.94)*****	
Deviation in Worked on problem, then revised solution	**0.17(2.10)**	
Use of additional Instructional Resources		
Review Session	-2.66(1.66)	-0.65(1.29)
Office Hours	3.39(1.82)	0.07(1.26)
Piazza (Reference: Did not use Piazza)		
Used Piazza passively	-0.99(1.81)	-2.28(1.52)
Used Piazza actively	3.07(2.24)	-2.68(1.86)
Use of written/online resources		
Memorized key words, concepts, molecules, and reactions	-3.67(1.93)*	-0.69(1.45)
Reviewed problem types that might appear on quiz/exam	-2.77(1.83)	-0.62(1.36)
Tested understanding of material	-1.71(1.80)	1.33(1.33)
Built connections among ideas	-0.64(1.93)	
Average of Built connections among ideas		**4.17(2.79)**
Deviation in Built connections among ideas		**-3.07(1.60)***
Identified underlying concepts	-1.05(1.93)	1.36(1.47)
Reviewed problem-solving approaches	-2.07(1.87)	0.36(1.45)
Use of practice problems		
Memorized key words, concepts, molecules, and reactions	1.91(1.67)	0.04(1.31)
Reviewed problem types that might appear on quiz/exam	2.44(1.68)	0.30(1.22)
Tested understanding of material		1.81(1.21)
Average of Tested understanding of material	**10.63(3.04)*****	
Deviation in Tested understanding of material	**-4.13(1.72)****	
Built connections among ideas		-0.72(1.66)
Average of Built connections among ideas	**4.86(4.21)**	
Deviation in Built connections among ideas	**-4.43(2.25)****	
Identified underlying concepts		0.39(1.17)
Average in Identified underlying concepts	**7.18(2.80)****	
Deviation in Identifying underlying concepts	**-1.65(1.89)**	

Table 6.4 Continued

Fixed and Random Effects for the HLM Models 2 & 4 (Bold corresponds to new variables)

	Model 2 PTP Score (100 total points)	Model 4 Mechanism Score (100 total points)
Use of practice problems – Continued		
Reviewed problem-solving approaches	1.91(1.58)	
Average in Reviewed problem-solving approaches		**2.04(2.32)**
Deviation in Reviewed problem-solving approaches		**-3.11(1.41)****
Group Study (Reference: Did not engage in group work)		
Average of Passively engaged in groupwork	**-8.42(5.13)**	**-16.88(4.64)*****
Deviation in Passively engaged in groupwork	**4.89(3.45)**	**5.20(2.27)****
Actively engaged in groupwork	-1.72(1.74)	-2.06(1.40)
Random Part		
Between student variance ψ	296.52	293.77
Within student variance θ	294.67	142.73

*-p-value approaching significance **p-value<0.05 ***p-value<0.001

Model 2 was then used to further probe the association between time spent studying and PTP scores (See Table 6.4). This analysis showed that students who always studied 7-10 hours the week before the exam were not estimated to perform significantly higher than those who always studied less than four hours. However, students who changed from studying less than four hours to studying 7-10 hours were estimated to on average perform 5.73 points higher on PTP questions, although this result was only approaching significance. Students who always studied more than 10 hours before an exam were also not estimated to perform significantly higher than those who always studied less than four hours. However, students who changed from studying less than 4 hours to studying more than 10 hours, were estimated to on average perform 6.92 points higher when they studied more.

In-Class study behaviors

Model 1 indicated, after controlling for all other covariates, that working on a problem, then revising the solution when the instructor went over the problem was associated with higher scores on PTP questions (See Table 6.3). Students who did so were estimated to on average score 3.49 points higher than those who did not. Model 3 indicated however that for mechanism questions only collaborating/talking with neighbors during lecture was associated with performance. Students that collaborated/talked with neighbors were estimated to on average perform 4.71 points higher on mechanism questions than students who did not (See Table 6.3).

Model 2 indicated that, controlling for all other covariates, students who always worked on a problem and then revised the solution when the instructor went over it were estimated to on average perform 8.80 points higher on PTP questions than those who never did so (See Table 6.4). However, students who worked on a problem and then revised the solution on one occasion but not another did not perform significantly differently on the two occasions. Model 4 indicated that for mechanism questions students who always talked/collaborated with their neighbors during lecture were estimated to on average score 6.83 points higher than those who never did so (See Table 6.4). Students who talked/collaborated with neighbors on one occasion were estimated to score 3.96 points higher than on an occasion where they did not talk/collaborate with their neighbors.

Use of additional instructional resources

Model 1 indicated that, students' PTP scores were not significantly associated with their use of additional instructional resources. Students who attended review sessions were estimated to on average score 2.76 points lower than student who did not, though this association was only approaching significance. Students who attended office hours in contrast were estimated to on average score 3.31 points higher than students who did not, though this association was also only approaching significance. Student's mechanism scores were also not significantly associated with their use of additional resources.

Use of online/written resources

Students' PTP scores were not significantly associated with their use of online or written resources, controlling for all other covariates (See Table 6.3). However, students' mechanism scores were significantly associated with whether they predominantly built connections between ideas while reviewing online/written resources. Model 3 indicated that, controlling for all other covariates, students who built connections between ideas were estimated to score 3.39 points higher on mechanism questions than those who did not.

Interestingly Model 4 showed that that neither between-student differences nor within-student differences were associated with student scores on Mechanism questions. Students who built connections between ideas on a given occasion were estimated to score 3.07 points lower than on an occasion where they did not use this strategy but this association was only approaching significance.

Use of practice problems

Model 1 indicated that various uses of practice problems were associated with performance on PTP questions. Students who predominantly used practice problems to test their understanding of the materials were estimated to on average score 6.08 points higher on PTP questions than students who did not, controlling for all other covariates (See Table 6.3). Students who predominantly used practice problems to build connections among ideas were estimated to, on average, score 5.00 points higher than those who did not. Finally, students who predominantly identified underlying concepts were estimated to score 3.39 points higher than students who did not, controlling for all over covariates. On the other hand, mechanism scores were only associated with whether students predominantly reviewed problem-solving approaches while working on practice problems. Model 3 indicated that students who predominantly reviewed problem-solving approaches were estimated to, on average, score 3.00 points higher controlling for all other covariates (See Table 6.3).

Model 2 indicated that students who always used practice problems to test their understanding were estimated to on average perform 10.63 points higher than students who never used practice problems to test their understanding (See Table 6.4). However, students who changed in their usage of this strategy were estimated to on average score 4.13 points lower when they used it than when they did not. Students who always built connections among ideas while working on practice problems did not perform significantly differently than students who never built connections among ideas. However, students who built connections between ideas on a given occasion were estimated to score 4.43 points lower than on an occasion where they did not use this strategy. Students who always used practice problems to identify underlying concepts were estimated to on average score 7.18 points higher than students who never did so, controlling for all covariates. Changes in usage of this strategy were not significantly associated with PTP scores.

Model 4 indicated that students who always used practice problems to review problem-solving approaches did not perform significantly differently on mechanism questions that those who never used this strategy, controlling for all other covariates (See Table 6.4). However, students who changed in their usage of this strategy were estimated to on average score 3.11 points lower when they used it than when they did not.

Use of group work

Model 1 indicated that students who passively engaged in group work were estimated to score lower on both PTP and mechanism questions than those who did not engage in group work, controlling for all other covariates (See Table 6.3). Students who actively engaged in group work did not perform significantly differently than those who did not engage in groupwork at all, but did perform on average an estimated 1.97 points higher on mechanism questions than students who passively engaged in group work. They also scored 2.75 points higher on PTP questions, though this result was only approaching significance.

Model 2 then indicated however that, neither subject means nor occasion-specific deviations were associated with students PTP scores. (See Table 6.4). Model 4 however indicated that, controlling for all other covariates, students who always engaged in passive group work were estimated to on average score 16.88 points lower on mechanism questions than students who did not engage in group work. In contrast though student who changed in their form of engagement were estimated to on average score 5.20 points higher when they passively engaged in group work than when they did not engage in group work.

Data Analysis

This chapter sought to explore the relationship between study behaviors and student outcomes on different categories of organic chemistry questions, specifically predict the product and mechanism questions. It also explored the similarities and differences in study behaviors associated with these two common question categories, and how they compared to those associated with overall exam performance in organic chemistry.

Predict-the-Production Questions

Students' scores on Predict the Problem type questions were overall associated with a wide-range of study behaviors. Both studying 7-10 hours and more than 10 hours were associated with higher scores than studying less than four hours. The observation that subject means for both time durations were not associated with student exam scores, while occasion-specific deviations were, suggests that consistently studying more than 7 hours the week before the exam is not associated with higher PTP scores. This may be due to the fact that this variable only looks at study hours the week before the exam and some of the students studying less than four hours may be studying regularly week-to-week while some of the students studying 7-10 hours and more than 10 hours the week before the exam may be engaging in cramming behavior. That deviations in studying less than four hours and studying both 7-10 and more than 10 hours are associated with higher outcomes suggests that increasing one's study time the week before an exam may be useful to students who want to improve their performance on PTP questions.

It was notable that PTP questions were particularly associated with students working on practice problems. This was true both in the classroom and outside the classroom. In-class students who worked on problem, and then revised them after the instructor went over them were estimated to on average perform higher on PTP question than those who did not. However, when subject-mean center variables were used results indicated that subject means were associated with higher PTP scores but occasion-specific deviations were not. This can potentially be interpreted in one of two ways. Either consistent usage of the study behavior is important and simply using it on one occasion is insufficient or students who have a stronger understanding of the material are better able to or more likely to engage in this behavior. The second may be a particularly plausible explanation as in-class observations indicated that students usually only had a few minutes to work on problems.

Students usage of practice problems was also associated with student PTP scores. Students who predominantly tested their understanding of the material, built connections among ideas, and identified underlying concepts were all estimated to have higher scores on PTP questions. It is interesting that these are in many ways more active behaviors than the other three in which students predominantly memorized key words, concepts, molecules, and reactions, reviewed problem types that might appear on quiz/exam, and reviewed problem-solving approaches.

The results of model 2 however showed some unusual results. Students who always tested their understanding of the material when doing practice problems were estimated to score significantly higher than students who never did so. However, when looking at occasion-specific deviations, students who tested their understanding when working on practice problems before an exam scored 4.13 points lower than on occasions where they did not test their understanding. This overall seems to indicate that, while consistent usage of this strategy is associated with higher performance, addition of this strategy may actually worsen student scores. This is surprising as metacognitive strategies such as self-testing are often associated with higher performance (Entwistle & McCune, 2004). However, it is possible that when students begin testing their understanding, they may not be very effective at doing so and it is only with time and practice that they develop the ability to accurately test their understanding.

Interestingly subject means or occasion-specific deviations for building connections among ideas were not significantly associated with PTP scores. However, students who changed in whether they used this strategy were estimated to score 4.43 points lower on occasions where they built connections among ideas when working on practice problems than on occasions where they did not. These results may be due to the relatively small sample of students who used study behavior, and future work, possibly pooling multiple semester of Organic Chemistry 2, may be needed to determine how average usage and differences in usage are associated with PTP scores. However, the negative association with occasion-specific deviation may indicate that when students initially begin trying to build connections among ideas they are not very effective at this practice and it is only with consistent practice that they are able to build deeper connections between ideas.

Students who always identified underlying concepts were estimated to on average score 7.18 points higher than students who never did so. However, occasion-specific deviations were not associated with PTP scores. This pattern has been seen previously and again may suggest either that stronger students are more likely or better able to engage in this study behavior or that consistent usage of it is important. It's also possible that students are not immediately effective at identifying concepts and that it takes time to develop the ability to do so effectively.

Finally, passive engagement in group work was associated with lower scores on PTP questions. However, students who always used this strategy were not estimated to score significantly different than students who never worked in groups. There was also no association between occasion-specific deviations and PTP scores. This is likely due to the small sample size and future work may need to pool multiple semester of Organic Chemistry 2. The results from Model 1 however would suggest that this strategy is more common with students who are struggling with PTP questions or that usage of this strategy is unhelpful and leads to lower scores.

Mechanism Questions

Students' scores on Mechanism questions, in comparison with PTP questions, were associated with fewer behaviors. They were specifically associated with Collaborating/Talking with neighbors during lecture, building connections among ideas while reviewing online/written resources, reviewing problem solving approaches while working on practice problems, and passive engagement in groupwork.

Mechanism questions were not associated with time spent studying. This may suggest that time spent studying is overall less significant or that in order to see gains students need to study much more than 10 hours the week before the exam. This is very different from PTP questions were studying more than seven hours is associated with higher exam scores.

In contrast with PTP questions, Mechanism questions were positively associated with collaborating/talking with neighbors during lecture. Interestingly both average usage of the strategies and occasion-specific deviations were associated with higher Mechanism Scores. This would seem to strongly indicate that collaborating while in class is helpful to students on mechanism questions and that addition of this study behavior can lead to gains on mechanism scores.

Mechanism sub-scores were also associated with use of online/written resources. Specifically, students who built connections among ideas while reviewing online/written resources were estimated to on average score higher than those who did not. However, in Model 4 neither within-student or between-student effects were significant. This is likely due to this study behavior being fairly uncommon (See Appendix D) and future work with more students may help probe whether deviations in usage are associated with outcomes.

Mechanism questions were also associated with the use of practice problems. Specifically, students who reviewing problem-solving approaches were estimated to on average score higher than students who did not. Interestingly after applying subject-mean centered covariates only occasion-specific deviations were associated with exam scores. What makes this particularly unusual is that this analysis indicated that students on average perform worse on occasions where they review problem solving approaches than on occasion when they do not. This may suggest that when students begin to use this study behavior, they are not very effective at recognizing helpful problem-solving approaches. The fact that consistent usage of this study behavior is not associated with higher outcomes may suggest that students do not improve at this study behavior with increased usage or that frequent usage of this strategy is not characteristic of higher performing students. Since this survey did not ask students to specify what they meant by problem-solving approaches, future work would be needed to clarify what students mean when they say that they reviewed problem solving approaches and to determine if this is consistent for students who receive low and high scores on mechanism questions.

Finally, score on mechanism questions, like PTP questions, are negatively associated with passive engagement in groupwork. While Model 4 showed that students who always engage in passive group on average were estimated to score 19.30 points lower than students who never engage in groupwork, students who changed between not working in groups and passively engaging in groups scored higher on exams when their studying involved passive group work. This suggests that students who are struggling are more likely to engage in passive groupwork but that, at least for mechanism questions, engagement passive groupwork is more helpful than not working in a group. Collaborative work seems to be particularly helpful on mechanism questions, in a way that is not seen in PTP questions.

Comparison

The most notable result of this analysis is that there is almost no overlap in the study behaviors associated with performance on predict-the-product questions and mechanism questions (See Table 6.3). This strongly indicates that treating exams as simple outcomes in research may ignore important differences in questions. It also indicates that the preparation required for students to do well on each of these categories of questions is very different. In fact, the only study behavior significantly associated with student performance on both mechanism and PTP sub-scores was passive engagement in group work, which was negatively associated with both.

Otherwise the differences between PTP and mechanism questions were quite dramatic. PTP questions were associated with students studying 7-10 hours and 10+ hours, while mechanism questions were not associated with study time. When considering in-class study behaviors, performance on mechanism questions was associated with collaboration with peers, while performance on PTP question was associated with attempting problems in class and revising them. Attending office hours was associated with higher PTP scores, but not higher mechanism scores and only mechanism scores were associated with reviewing online/written resources. Finally, while both PTP and Mechanism questions were associated with practice problem use, they were each associated with different uses of practice problems. Broadly mechanism questions seemed to be more associated with collaborative work while PTP questions seemed to be more associated with practice problems. It is possible that since mechanism questions often ask students to explain fairly complex systems, collaborative situations are particularly conducive to students learning. PTP questions, in contrast, ask students to develop predictions and be familiar with reactions, which may best be supported by active and reflective use of practice problems.

These results were also different from those seen for overall exam performance. Overall exam scores were associated with studying more than 10 hours, writing down one's own thoughts/ideas when taking notes, and testing one's understanding when doing practice problems (See Table 5.2). Study behaviors associated with PTP questions are most similar to overall exam outcomes, sharing three out of four behaviors. However, PTP questions are also associated with many other study behaviors that are not significantly associated with overall exam outcomes. Mechanism questions, in contrast, were associated with very different strategies than overall exam scores, only sharing that passive engagement in group work was negatively associated with performance. Interestingly mechanism sub-scores, PTP scores, and overall exam score were all associated with different in-class study behaviors. This may suggest that all these behaviors are helpful but that they might be helpful on different categories of questions. In only looking at only overall exam scores we may miss potentially important study behaviors that are associated with particular categories of questions.

Conclusions

Overall this analysis indicated that different categories of questions are associated with very different study behaviors. There are also notable differences in the study behaviors associated with mechanism and predict the product sub-scores and overall exam performance. These results are not fully generalizable; the study behaviors instrument is by definition situated in the context of Organic Chemistry 2. However, the stark differences suggest that until we consider the make-up of items within exams, we are not seeing the full picture of how student study behaviors can influence student performance. These differences can have implications for both the ways that instructors design assessments in organic chemistry and the ways that they can advise students.

In terms of advising, these results indicate that student performance on PTP and mechanism questions are influenced by very different study behaviors. In order to better advise students, instructors should consider what kinds of questions students are struggling with and direct them to resources that would help them with those types of questions. For example, the analysis conducted in Chapter 4 seemed to indicate that collaborating with neighbors during lecture wasn't associated with higher outcomes. However, if a student is struggling with mechanism questions results from this chapter would seem to indicate that addition of this study behavior would be helpful. On the other hand, if a student was struggling with predict-the-product questions this strategy does not appear to have any effect.

In terms of exam design, this suggest that instructors should take into account the types of questions that they use in their exams and be aware that the structure of these questions is influential, not just the concepts that the question measures. It may also be useful for instructors to consider the major categories of questions that make up their exams and check that they are associated with study behaviors that lead to deeper learning. A problem type, for example, positively associated with study behaviors emphasizing memorization or passive engagements will not effectively measure student learning. It may also encourage students to learn less effectively, particularly in non-major courses where students are not intrinsically motivated to learn the material.

While these results were notable it is worth discussing some of the limitations of this work. Some of these limitations were previously discussed in chapter 4. The most notable in this analysis however may be the small sample size on some variables, which seemed to be most influential when introducing subject-mean centered covariates. Follow-up work with a larger sample size may also be called for since there were a large number of study behaviors being tested, particularly after introducing subject means and occasion-specific deviations for significant study behaviors. Finally, while this work indicates that categories of questions are associated with different study behaviors, even within a category of questions, items can have different relationships to the curriculum. Whether a given question is a problem-type of exercise-type question may be important, and whether a category of questions is more likely to be made of up of exercises or problems may also be important. The relationship between study behaviors, problems, and exercises will be explored in more detail in the next chapter.

Chapter 7: Problems, Exercises, and Study Behaviors

Introduction and Research Questions

Chapter 6 explored what types of study behaviors were associated with student scores on different types of questions. It also described how different categories of questions, namely mechanism and PTP questions, are associated with very different study behaviors. These results emphasize the importance of considering the structure of assessments when considering what students, and instructors, can do to improve student performance. However, what was not considered in the previous chapter was the relationship that a given question has to the curriculum. This relationship is incredibly important, since the difficulty of a question often depends on it and it is possible that different study behaviors and reasoning strategies (Chapter 8) may impact student performance on questions differently depending on their relationship to the curriculum.

As described in Chapter 2, organic chemistry problem-solving literature has considered it important to acknowledge the differences between problems and exercises (Bodner & Herron, 2002). This has often been used to emphasize how questions often seem like exercises to experts and problems to students. Many studies have also chosen to focus on more challenging questions so as to ensure that they are studying problem-solving and not just students' work on exercises (Bodner, 1987; Flynn, 2014).

While one can try to develop problems that are inherently problems or exercises, it is impossible to determine whether a question acts as a problem or an exercise to a given student without probing their thinking(Bodner, 1987). However, one can however determine whether a given question is a problem or an exercise in relationship to the curriculum. In short whether a student, given the curriculum of the class, would be able to develop the algorithms needed to view a question as a rote exercise. Questions can therefore be described as **exercise-type** questions, which are very similar to examples and material previously seen in the course, or as **problem-type** questions, which involve synthesis of multiple ideas of applying concepts in a new context. Throughout this chapter I simply refer to these types of questions as exercises and problems for sake of simplicity. This does not however indicate that these questions are universally seen by students as problems or exercises. It is simply described the relationship of the question to the curriculum and the exposure that students have had to similar concepts, contexts, and examples. By developing a deeper understanding of the curriculum through classroom observation and analyses of practice problems one can determine the relationship that individual questions have to the curriculum and explore how student study behaviors are associated with both problems and exercises. This chapter is focused on exploring the nature of these associations. In particular it seeks to explore student performance on two questions with very different relationships to the curriculum and answer the following research questions:

1) How are students' study behaviors associated with performance on problem-type questions (specifically within the sub-category of the predict-the-product questions)?

2) How are students' study behaviors associated with performance on exercise-type questions (specifically within the sub-category of the predict-the-product questions)?
3) Is performance on problem-type and exercise-type questions correlated with different study behaviors?

Methods: Participants and Variables

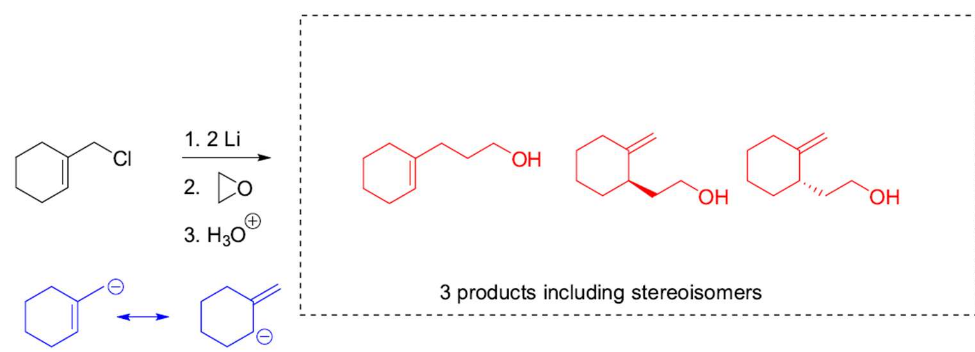
There were 357 students from Organic Chemistry 2 included in this analysis. This population consisted of students who had completed the first survey as well as the first midterm. This is equivalent to the participants at Occasion 1 in the previous two chapters. All independent variables used in this analysis were previously described in Chapter 3 in the Variable section. This section will describe the response variables used in this analysis.

Response Variables

In order to analyze the potential relationship between study behaviors and performance on both problems and exercises, two questions were selected from the first midterm; Q1 was characterized as a problem and Q2 as an exercise. Both questions were the same category of questions (predict-the-product) since Chapter 6 showed that different categories of questions are associated with different study behaviors. Student interviews also centered on predict-the-product questions meaning that further analysis could be conducted on the problem-solving approaches used on these questions and differences between problems and exercises (See Chapter 8). Only one question of each type was used in this analysis because PTP problems were fairly uncommon and no more than one was identified for any given exam.

In order to determine if a given question was a problem or an exercise, first the question was broken down into its problem-solving step (See Tables 8.2 and 8.3 for examples). Then examples in course materials such as: instructor lecture handouts, problem sets, quizzes, and practice exams were searched for related examples which were coded into two categories: **Highly Similar** and **Somewhat similar**. Classroom observations were also used, as needed, to provide additional information about how the questions were presented in class.

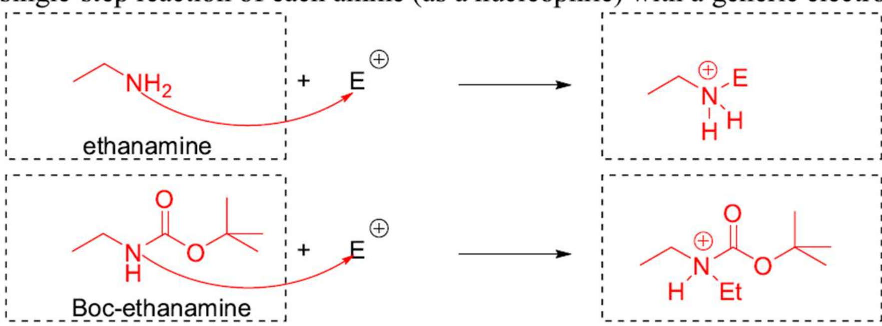
Figure 7.1 Highly Similar example to Question 2



Highly similar examples were those where at least half the problem-solving steps matched. For example, in Figure 6.1 the example given is highly similar to question 1 because the problems match in almost every problem-solving step. The only difference is that in Question 2 the alkyl lithium species reacts with a carbonyl and in the highly similar example it reacts with an epoxide. **Somewhat similar** examples were those where less than half the problem-solving steps matched. For example, in Figure 6.2 this example is considered somewhat similar to Question 1 because it is only relevant to one step in the problem-solving process.

Figure 7.2 Somewhat Similar example to Question 1

b. Draw ethanamine and Boc-ethanamine, then add curved arrows and the resulting products for a single-step reaction of each amine (as a nucleophile) with a generic electrophile, E+.

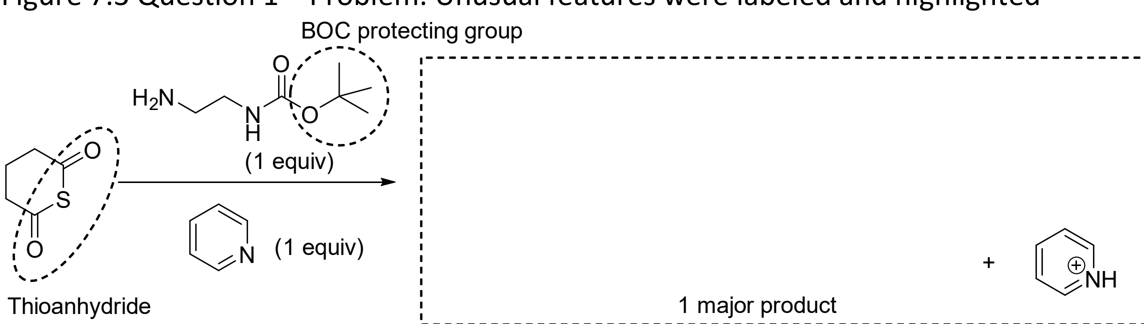


c. Explain why Boc-ethanamine is much less nucleophilic than ethanamine.
If Boc-ethanamine reacts with an electrophile, it destroys the conjugation (resonance stabilization) of the nitrogen lone pair with the carbonyl that is present in the starting structure. There is no similar stabilization in ethanamine.

A question was considered an exercise if there were at least three highly similar examples present in the curriculum and if each step in the problem-solving process had been seen in at least three examples (either highly similar or somewhat similar examples). This meant that students would have had ample exposure to all the parts of the question and be comfortable seeing these components together. If either rule was not satisfied the question was considered a problem. This categorization was then verified with the instructor of Organic Chemistry 2, to ensure that that the coding scheme and categories were in line with her intentions for the question.

PTP_Q1_Problem: Continuous response variable corresponding to a student's score on a predict-the-product problem asking students to predict the products of an acyl substitution reaction under basic conditions (See Figure 7.1). Students could score between 0 and 6 points.

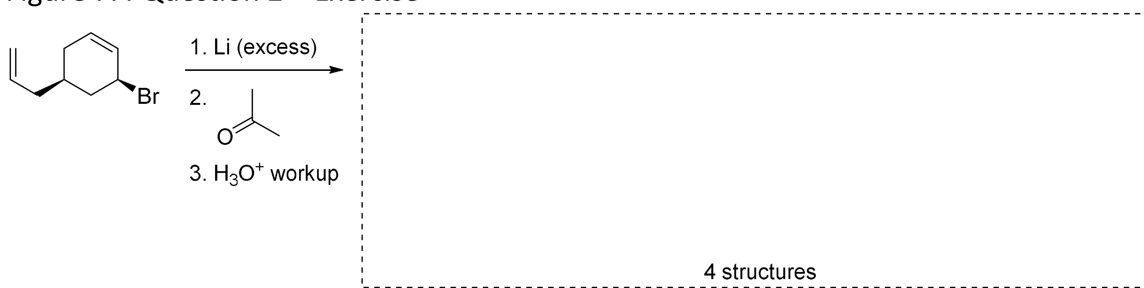
Figure 7.3 Question 1 – Problem. Unusual features were labeled and highlighted



This question was identified as a problem because students were required to synthesize various concepts that were presented in-class and on homework assignments, there were no highly-similar examples. This meant that it included components that were unusual for an acyl substitution type problem, a BOC protecting group which students had only seen in the context of a different reaction and a thioanhydride functional group which had only shown up once in a practice exam (See Figure 7.1 and Table 8.2 for more information).

PTP_Q2_Exercise: Continuous response variable corresponding to a student's score on PTP exercise asking students to predict the products of an allylic carbanion and a ketone (See Figure 7.2 and Table 8.3). Students could score between 0 and 6 points.

Figure 7.4 Question 2 – Exercise



This question was identified as an exercise because there were multiple highly similar examples seen in class and on their homework assignments (See Figure 7.1). Almost identical problems had been seen several times and there were many somewhat similar problems present as well. Reactions between allylic carbanions and ketones were also explicitly discussed during lecture.

Methods: Statistical Analysis

Two multiple linear regression models were used. The first regressed PTP_Q1_Problem onto Study Hour variables, In-Class Study Behavior variables, Additional Instructional Resources variables, Use of written/online resources variables, Use of practice problem variable, and Group study variables. All 24 variables were tested for collinearity; Variance Inflation favors (VIF) were all less than 4 for all variables indicating that there was no collinearity. The studentized deleted residuals were not normal and variance did not appear constant so robust standard errors were used (See Appendix D).

Model 1: Problem-type Question

$$PTP\_Q1\_Problem = \beta_0 + \beta_1(study\_hours\_4\text{-}7hrs)_i + \beta_2(study\_hours\_7\text{-}10hrs)_i + \beta_3(study\_hours\_10\text{+}hr)_i + \beta_4(lecture\_collaboration)_i + \beta_5(worksheet\_revision)_i + \beta_6(notes\_ownswords)_i + \beta_7(review\_session)_i + \beta_8(office\_hours)_i + \beta_9(piazza\_passive)_i + \beta_{10}(piazza\_active)_i + \beta_{11}(review\_memorize)_i + \beta_{12}(review\_Ptypes)_i + \beta_{13}(review\_testunderstanding)_i + \beta_{14}(review\_connectideas)_i + \beta_{15}(review\_concepts)_i + \beta_{16}(review\_Psolving)_i + \beta_{17}(problem\_memorize)_i + \beta_{18}(problem\_Ptypes)_i + \beta_{19}(problem\_testunderstanding)_i + \beta_{20}(problem\_connectideas)_i + \beta_{21}(problem\_concepts)_i + \beta_{22}(problem\_Psolving)_i + \beta_{23}(grpstudy\_passive)_i + \beta_{24}(grpstudy\_active)_i + \varepsilon_i$$

The second model regressed PTP_Q2_Exercise onto Study Hour variables, In-Class Study Behavior variables, Additional Instructional Resources variables, Use of written/online resources variables, Use of practice problem variable, and Group study variables. All 24 variables were tested for collinearity; VIF were less than 4 for all variables indicating that there was no collinearity. The studentized deleted residuals were not normal and variance did not appear constant so robust standard errors were used (See Appendix D).

Model 2: Exercise-Type Question

$$PTP\_Q1\_Exercise_i = \beta_0 + \beta_1(study\_hours\_4\text{-}7hrs)_i + \beta_2(study\_hours\_7\text{-}10hrs)_i + \beta_3(study\_hours\_10\text{+}hr)_i + \beta_4(lecture\_collaboration)_i + \beta_5(worksheet\_revision)_i + \beta_6(notes\_ownswords)_i + \beta_7(review\_session)_i + \beta_8(office\_hours)_i + \beta_9(piazza\_passive)_i + \beta_{10}(piazza\_active)_i + \beta_{11}(review\_memorize)_i + \beta_{12}(review\_Ptypes)_i + \beta_{13}(review\_testunderstanding)_i + \beta_{14}(review\_connectideas)_i + \beta_{15}(review\_concepts)_i + \beta_{16}(review\_Psolving)_i + \beta_{17}(problem\_memorize)_i + \beta_{18}(problem\_Ptypes)_i + \beta_{19}(problem\_testunderstanding)_i + \beta_{20}(problem\_connectideas)_i + \beta_{21}(problem\_concepts)_i + \beta_{22}(problem\_Psolving)_i + \beta_{23}(grpstudy\_passive)_i + \beta_{24}(grpstudy\_active)_i + \varepsilon_i$$

Results: Descriptive Statistics

While 538 students were registered in Organic Chemistry 2, only 66% of students were included as participants in analysis (N = 357). 383 individuals completed the first survey; however, participants were dropped from the analysis if they had no reported exam scores as this indicated either a very early drop of the course or that they had mis-reported their student ID in the survey. Participants were also dropped they were missing independent variables.

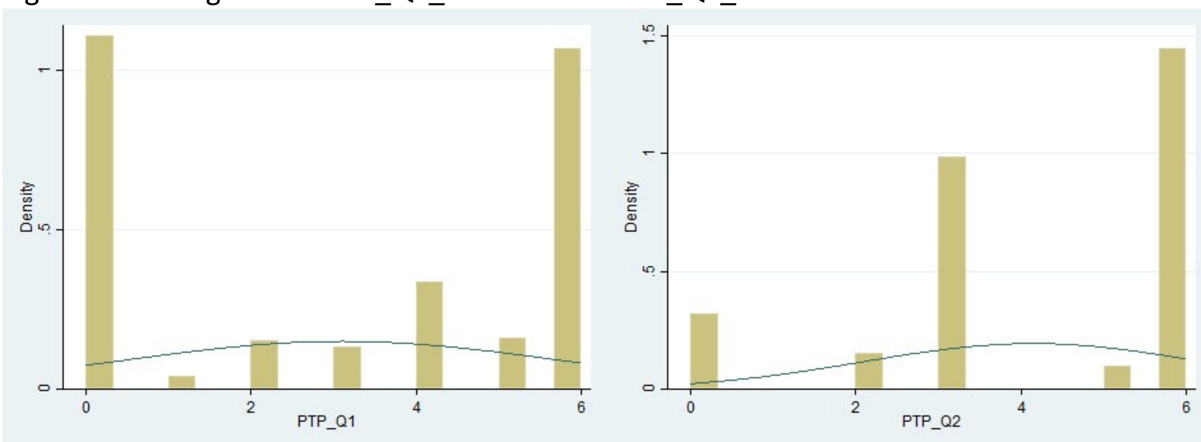
Table 7.1

Summary of Continuous Response Variables

Variable	Mean	Standard Deviation	Minimum	Maximum
PTP_Q1_Problem	3.10	2.64	0	6
PTP_Q2_Exercise	4.14	2.05	0	6

Students on average scored higher on the exercise than on the problem. An unpaired t test indicated that this difference was significant ($t(357) = 1.04$, $p<0.001$).

Figure 7.3 Histograms of PTP_Q1_Problem and PTP_Q2_Exercise



Histograms of student scores indicated that neither distribution was normal. PTP_Q1 has higher distributions at the low and high ends and PTP_Q2 is negatively skewed (See Figure 7.3). Studentized deleted residuals were also not normal so robust standard errors were used (See Appendix D). Other modifications to the response variables were considered, but in the end not adopted since this format allowed for easier comparison with results from previous chapters. The assumption of constant variance is also broken which supports the use of robust standard errors (See Appendix D).

Table 7.2

Comparison of Student Scores on PTP_Q1_Problem and PTP_Q2_Exercise

	Number of Participants	Percentage of Participants
Higher score on PTP_Q1_Problem	85	24%
Higher score on PTP_Q2_Exercise	164	46%
Same score on both problems	108	30%

While participants on average scored higher on the exercise than on the problem, 24% (*N* =85) scored higher on the problem (See Table 7.2). From the frequency table of categorical variables (Appendix D) we can see that prior to the first midterm 54% of participants studied more than 10 hours (*N* = 198). The majority of participants actively used in-class study behaviors. Only 23% of students attended review sessions (*N* = 81) and 20% attended office hours (*N* = 71). Most participants used the Piazza message boards though only 22% used it actively (*N* = 79). The most common usage of online/written resources was reviewing problem types that might appear on a quiz/exam which was used by 72% of participants (*N* = 257). The least common was building connections among ideas which only 24% of participants used (*N* = 85). The same was true for participants' usage of practice problems. 68% of participants reviewed problem types that might appear on a quiz/exam (*N* =244) and 14% of participants built connections among ideas (*N* = 50). Most students participated in some type of selforganized group work outside of class. 8% engaged in passive group (*N* = 28) and 64% engaged in active group work (*N* = 230).

Results: Statistical Analysis

Two multiple linear regression models were used to answer the research questions. These models regressed student scores on exam questions onto students' usage of study behaviors to determine how students' scores on a problem-type question and an exercise-type question were associated with student study behaviors.

Table 7.3

Study Behaviors and Scores on Problem-Type and Exercise-Type Questions

	Model 1 Question 1 – Problem (6)	Model 2 Question 2 – Exercise (6)
Study Hours (Reference: <4 hours)		
4-7 hours	-0.09(0.61)	0.50(0.47)
7-10 hours	0.43(0.53)	0.99(0.43)**
10 + hours	0.88(0.49)*	0.68(0.40)*
In-Class Study Behaviors		
Collaborated/Talked with neighbors during lecture	-0.02(0.35)	-0.42(0.26)
Wrote down own thoughts/words when taking notes	0.25(0.30)	0.62(0.01)**
Worked on problem, then revised the solution when the instructor went over the problem	0.93(0.35)***	0.12(0.27)
Use of additional Instructional Resources		
Review Session	-1.02(0.33)***	-0.21(0.27)
Office Hours	0.06(0.35)	-0.45(0.28)
Piazza (Reference: Did not use Piazza)		
Used Piazza passively	0.45(0.35)	0.09(0.26)
Used Piazza actively	0.53(0.40)	0.03(0.33)
Use of written/online resources		
Memorized key words, concepts, molecules, and reactions	-0.27(0.38)	-0.04(0.33)
Reviewed problem types that might appear on quiz/exam	-0.21(0.41)	-0.26(0.30)
Tested understanding of material	0.19(0.38)	0.58(0.30)*
Built connections among ideas	-0.52(0.43)	0.10(0.35)
Identified underlying concepts	0.17(0.41)	-0.21(0.33)
Reviewed problem-solving approaches	0.01(0.38)	0.02(0.32)
Use of practice problems		
Memorized key words, concepts, molecules, and reactions	-0.03(0.35)	-0.34(0.29)
Reviewed problem types that might appear on quiz/exam	0.28(0.34)	0.42(0.26)
Tested understanding of material	0.61(0.33)*	0.24(0.27)
Built connections among ideas	0.88(0.47)*	0.18(0.33)
Identified underlying concepts	0.50(0.35)	0.52(0.28)*
Reviewed problem-solving approaches	0.01(0.34)	0.23(0.27)
Group Study (Reference: Did not engage in group work)		
Passively engaged in groupwork	-0.83(0.58)	-0.18(0.44)
Actively engaged in groupwork	-0.59(0.32)*	-0.08(0.27)

*-p-value approaching significance **p-value<0.05 ***p-value<0.001

Study Hours

Model 1 indicated that, controlling for all other study behaviors, students who studied more than 10 hours the week before the exam were estimated to on average perform 0.88 points higher than students who studied less than 4 hours on the problem, however this result was only approaching significance (See Table 7.3). Interestingly students who studied more than 10 hours were estimated to score significantly higher than students who studied 4-7 hours. Students who studied more than 10 hours were estimated to on average score 0.97 points higher than students studied 4-7 hours ($p<0.05$). Students who studied 7-10 hours, 4-7 hours, and less than four hours however did not perform significantly differently from each other. The set of dummy variables representing hours spent studying the week before the exam were found to be approaching significance at the 5% level ($F(3,332)=2.23$, $p=0.08$).

Model 2 showed that, controlling for all other covariates, students who studied 7-10 hours performed significantly higher on the exercise-type question condition than students who studied less than four hours (See Table 7.3). Students who studied more than 10 hours scored 0.68 points higher than those who studied less than four hours but this was only approaching significance. No other differences in study hours were significant. The set of dummy variables representing hours spent studying the week before the exam were found to not be significantly difference ($F(3,332)=1.92$, $p=0.13$).

In-Class Study Behaviors

Model 1 indicated that, controlling for all other covariates, only working on problems and then revising the solution when the instructor went over the problem the week before they exam was correlated with higher performance on the problem. Students who engaged in this behavior were estimated to score on average 0.93 points higher than students who did not (See Table 7.3).

Model 2 indicated that, in contrast, only writing down one's own thoughts and words while taking notes was correlated with higher performance on the exercise, controlling for all other covariates. Students who engaged in this behavior were estimated to on average score 0.62 points higher than students who did not do so.

Use of additional instructional resources

Model 1 indicated that, controlling for all other covariates, only attending review sessions the week before the exam was associated with student performance on the problem (See Table 7.3). Students who did so were estimated to on average score 1.02 points lower than students who did not do so. No uses of additional instructional resources were significantly correlated with students' scores on the exercise-type question.

Use of written/online resources

Neither students' scores on the problem-type (Question 1) nor the exercise-type question (Question 2) were significantly associated with their usage of written or online resources. Model 2 indicated, that controlling for all other covariates students who predominantly tested their understanding when reviewing written/online resources performed 0.58 points higher than those who did not but this result was only approaching significance.

Use of practice problems

Neither students' scores on the problem-type (Question 1) nor the exercise-type question (Question 2) were significantly associated with their usage of practice problems. Model 1 indicated that, controlling for all other covariates, students who predominantly tested their understanding of the material were estimated to on average score 0.61 points higher than those who did not, though this result was only approaching significance. Students who built connections among ideas while working on practice problems, were also estimated to score higher on question 1 though this result was also approaching significance (See Table 7.3). Model 2 indicated that, controlling for all other covariates, students who predominantly identified underlying concepts were estimated to on average score 0.52 points higher though these results were only approaching significance.

Group study

Neither students' scores on the problem (Question 1) or on the exercise (Question 2) were significantly correlated with their engagement in group work. However, students who engaged in active group work were estimated to score 0.59 points lower on question 1 than those who did not engage group work at all, however this result was only approaching significance.

Data Analysis

This chapter sought to determine what study behaviors were associated with a predict-the-product problem, what study behaviors were associated with a predict-the-product exercise, and what similarities and differences existed between these two questions. The most notable result was that the problem-type and exercise-type questions were each correlated with very different study behaviors. This is important because it indicates that relationship to the curriculum may not just vary the potential difficulty of a question but may require the use of different study behaviors.

For the problem, for example, students who studied more than 10 hours were found to perform significantly better than students who studied 4-7 hours. On the exercise question however, students who studied 7-10 hours scored significantly higher than students who studied <4 hours. This may suggest that students don't need to study as long to see higher scores on exercises as compared to problems. It may also be that different study hours may be associated with the development of different types of skills.

The two questions were also correlated with different in-class study behaviors. Students who wrote down their own thoughts/words when taking notes were estimated to on average score higher on the exercise. On the other hand, students who worked on problems and then revised the solution when the instructor went over the problem were estimated to on average score higher in the problem. Future work might need explore the nature of the thoughts/ideas that students are writing in their notes and why this might be associated with higher scores on exercises. Future work may also be needed to determine if working on problems in class and then revising helps students on problems or if it is simply a study behavior that students who are comfortable with problems are more likely to engage in.

Students who attended the review session the week before the first midterm were estimated to on average score 1.03 points lower than students who did not do so. The relationship however did not exist for the exercise-type question. This may suggest that either students who were struggling with problems were more likely to attend the review sessions or that the review sessions in Organic Chemistry 2 did not support students in developing the skills they needed to solve problems. The second interpretation is somewhat supported by observations which indicated that review sessions tended to be passive lectures which tried to simplify and categorize the material that students had been taught in lecture.

Interestingly neither use of online/written resources or practice problems was associated with students' scores on either problems or exercises. Some associations were found that were approaching significance which may indicate that this analysis might require a larger study population. This is supported by the results in the previous chapter that overall PTP questions are associated with students' usage of practice problems.

There were also notable differences between the study behaviors associated with the two questions analyzed in this chapter and study behaviors associated with overall PTP scores. Interestingly the problem-type question shared more study behaviors with overall PTP scores, with while the exercise condition had more differences. Future work will hopefully explore a larger number of PTP problems and exercises, to see if the study behaviors identified in these two problems can be generalized to PTP problems and exercises more broadly.

Conclusions

The results of this chapter are limited in their generalizability due to being situated within the context of Organic Chemistry 2. It was also limited by the use of only two questions to explore the differences between problems and exercises. Future work should consider whether these trends extend across a wider number of questions. This work should also explore if problems and exercises within other categories of questions (ie. Mechanism questions) have similar or different associations with study behaviors.

However even with these limitations, this chapter demonstrated the problems and exercises can be associated with very different strategies even within the same question type. This has implications for both how instructor advise students and design exams. Similar to the previous chapter, these results suggest that when advising students instructors should consider how students are doing on specific types of questions. For example, if a student in struggling on exercises, these results suggest that they should be encouraged to focus more on the way they write their notes and be reflective and metacognitive in developing them. In contrast, if students are struggling with problems, they should be encouraged to work on practice questions and be reflective and metacognitive while doing so.

These results also have implications for course and exam design. It was notable that attendance at the review session was negatively correlated with problem scores. This suggests that instructors should take care that the additional resources they provide support students on all types on questions. While Organic Chemistry 2 lectures were structured around active learning, the review sessions were not. The review session’s emphasis on summarizing key points may not have sufficiently supported students in developing the skills needed to solve problems. In terms of exam design, instructors should be aware that problems and exercises are not simply different in terms of difficulty, they also differentiate students based on the ways that they study.

Chapter 8: Problem-Solving in Problems and Exercises

Introduction

Chapter 7 showed that there were notable differences in the study behaviors associated with performance on problems and exercises in Organic Chemistry 2. In this chapter I expand this analysis to explore how students' decisions in problem solving influence their performance on problems and exercises in organic chemistry. This chapter specifically seeks to explore the reasoning strategies that students use to make decisions within a given question, how these differ between a problem and an exercise, and how students use of these strategies influence performance.

While several researchers in the field of chemistry education have explored student problem-solving, there has been little work directly comparing student problem solving on problem-type and exercise-type questions. Some work in the field of general chemistry has considered the differences between algorithm questions and conceptual questions (Cracolice, Deming, & Ehlert, 2008; Nakhleh & Mitchell, 1993; Zoller, Lubezky, Nakhleh, Tessier, & Dori, 1995). This work is limited in its applicability to organic chemistry. In general chemistry courses the distinction between algorithm and conceptual understanding questions is often more obvious. Many questions involve the application of mathematical formulas so it is often difficult to extricate the structure of the item from its relationship to the curriculum. Researchers in this work have therefore not usually not emphasized this distinction. Organic Chemistry questions, in contrast, require the use of non-mathematical problem solving so the difference between a problem and an exercise is entirely due to a questions' relationship to the curriculum (Cartrette & Bodner, 2010).

Much of the research on problem-solving in organic chemistry problem-solving, has intentionally focused on students' work on problems (Bowen & Bodner, 1991; Flynn, 2014). A few exceptions exist where researchers have considered the differences between problem-type and exercise-type questions. A recent example is Webber and Flynn's work which explored problem-solving in familiar and unfamiliar questions (Webber & Flynn, 2018). This paper indicated that there were differences in the problem-solving strategies that students needed to utilize to perform well on either type of question.

This chapter is interested on the decisions that students make within a given question and the reasoning strategies that students utilize to make those decisions and reach a solution. However, within research focusing on problem-type questions, researchers have tended to focus on the overarching reasoning strategy used by students to solve questions and have not addressed the variety of reasoning strategies that a student would use within a given problem to make decisions and reach a solution (Christian & Talanquer, 2012; Kraft, Strickland, & Bhattacharyya, 2010). A helpful idea for considering student decision-making and reasoning within a problem is the idea of a problem space. Problem space, also known as a search space is the idea that each problem consists of a series of possible answers and that the problem-solver must search within it for the possible answer(Klahr & Dunbar, 1988; Anderson, 2009). Problem-solvers must seek operations (or strategies) that allow them to navigate this space and will often seek to constrain the problem-space through information given in the problem. This chapter therefore seeks to answer the following research questions:

1) How do students make decisions when solving organic chemistry questions?
2) How do students' choices during problem solving exercises and problems?
3) What types of reasoning do they emphasize and value?

To answer the questions semi-structured think-aloud interviews were conducted. Analyses included: identifying the phases of problem-solving and reasoning strategies utilized; distinguishing between problem-solving for exercises and problems, and exploring the role of conceptual understanding in resolving tension points during problem-solving.

Methodology: Participants

Students in Organic Chemistry 2 were invited to participate in a 30-45 minutes semi-structured interview after their first midterm. During this interview students were asked about their study behaviors in Organic Chemistry 2 and then asked to think-aloud as they solved three Predict-the-Product (PTP) questions from the first midterm (See Chapter 3 and Appendix B for more information). After they had finished solving each question participants were asked to reflect on what they did to prepare for the question and what they could have done to prepare better. Participants were also asked follow-up questions about decisions they made during the think-aloud that were underexplained. Participants were then invited to complete follow-up interviews after the second midterm and around the time of the final exam. This study however will only discuss results from the first interview specifically the think-aloud portion.

After completing the interview each participant was assigned a pseudonym. While 22 students completed the first interview, only 15 were included in this analysis (See Table 8.1). Participants were excluded from the analysis if there were problems with the audio from their recording (such as students who requested that they not be audio-recorded) which limited my ability to create detailed verbatim transcripts. Three students were also excluded for insufficiently verbalizing their thoughts during the think-aloud process. Two participants were very nervous and did not verbalize their thinking as they solved the problem which meant that their interviews were unable to capture their think-aloud process. The third participant verbalized in a disjointed fashion and would describe features he noticed in the questions but did not express his rationale for this thoughts or decisions. Several students expressed nervousness as they solved question. Students were assured that I was only interested in their thinking and that they could stop the interview at any time. While students occasionally gave up on a question or took a break and came back to it, no student requested to end the interview.

Table 8.1

Interview Participants

Participant	Gender	Ethnicity	Organic Chemistry 1 Grade	Organic Chemistry 2 Grade	Question 1	Question 2
Tessa	Female	East Asian	A+	A+	Incorrect*	Correct
Neil	Male	White/Caucasian	A+	A	Incorrect	Correct
Andrew	Male	White/Caucasian	A	A+	Correct	Correct
Andrea	Female	South Asian; White/Caucasian	A-	A	Correct	Incorrect
Rebecca	Female	White/Caucasian	A-	A-	Correct	Correct
Patricia	Female	East Asian	A-	A+	Correct	Correct
William	Male	East Asian	B+	B+	Incorrect	Correct
Katie	Female	East Asian	B+	B	No response	Correct
Sophia	Female	Mexican American/Chicano; White/Caucasian	B+	A-	Correct	Correct
Amanda	Female	White/Caucasian	B+	A	Correct	Incorrect
Keira	Female	White/Caucasian	B+	B+	Incorrect	Correct
Nancy	Female	Filipino; White/Caucasian	B	A-	Incorrect	Correct
Olivia	Female	White/Caucasian	B-	C	Incorrect	Correct
Brianna	Female	Black/African American; White/Caucasian	C-	W	No response	No response
Samantha	Female	White/Caucasian	C-	C	Incorrect	Incorrect

As can be seen in Table 8.1 the interview participants had a wide-range of grades both in Organic Chemistry 2 and in the previous course in the series, Organic Chemistry 1. Students with A grades were overrepresented in the population and students with lower grades (B and lower) were underrepresented. It is notable however that several students who received Cs in the class were part of the study. One participant, Brianna, also chose to Withdraw from the class and retake in the next semester due to low grades on exams and quizzes. Female students were also overrepresented in the population as very few male students volunteered to participate in interviews. White students were also overrepresented in the population though there were volunteers from various ethnic groups.

Figure 8.1 Outcome of Question 1 – Problem

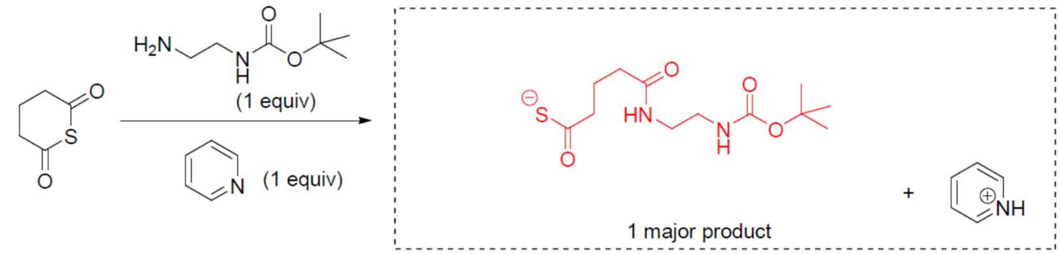


Table 8.2

Problem-Solving Process for Question 1: Problem

Representation	Problem-Solving
	Identify amine (circled) as reactive site. - Identify that amide N (in square) was not reactive because of conjugation or because of the BOC protecting group (in square).
	Identify carbonyl (circled) as reactive site - Identify that both were equivalent so you could react with either position
	Identify that amine will react with carbonyl, break the double bond, and form a tetrahedral intermediate.
OR	**Identify that pyridine acts as a base and will deprotonate the amine** - While it is more accurate to say that this reaction occurs after the amine reacts with the carbonyl, students who thought it happened first would have still gotten the correct answer.
	Identify that sulfur is the leaving group and that the carbonyl reforms - Recognize that the reaction is over and no other reactions will occur
	Possible distractors - BOC protecting group was usually seen with other classes of reactions - Reactions between O and S were seen in the class (though with different reagents)

Methodology: Questions

The analysis of this data began with the development of paraphrased interview transcripts. This preliminary analysis of the interviews indicated that two questions from the first interview were of particularly interest since successful student problem-solving (that which led to the correct answer) seemed very different. These two questions were from the first midterm exam and, as previously described in Chapter 7, had different relationships with the curriculum. They were identified to be a problem-type and an exercises-type question. The process for solving Question 1 – Problem can be seen in Table 8.2. This question involved an acyl substitution reaction under basic conditions, and while most of the steps identified had been seen previously by students they had not been seen in the same question. For example, while differences in amine and amide reactivity had been discussed previously, students had not been asked to pick between the two in previous PTP questions. Two aspects of the problem were also particularly unusual: the presence of a BOC protecting group (usually seen in peptide synthesis problems) and the presence of a thioanhydride functional group. Thioanhydride groups had only even been seen in one practice midterm, though related groups (thiols anhydrides) had been previously seen.

In contrast Question 2 – Exercise, in which an allylic carbanion reacted with a ketone, was very similar to examples students had seen previously. Both the formation of and reactions with carbanions (negatively charge carbons) had been previously discussed in lecture and multiple examples had appeared in lecture. The presence of resonance in these systems (a situation in which electron density is shared across multiple bond) had also been seen in both lectures and practice problems. This was also true about the formation of stereocenters (situations in which four different substituents are attached to one carbon meaning that multiple orientations are possible). Nothing about the problem was particularly unusual though students in interviews did occasionally have difficulty identifying the formation of a stereocenter and occasionally thought that the alkene side-chain could change orientation (See Table 8.3).

Figure 8.2 Outcome of Question 2 – Exercise

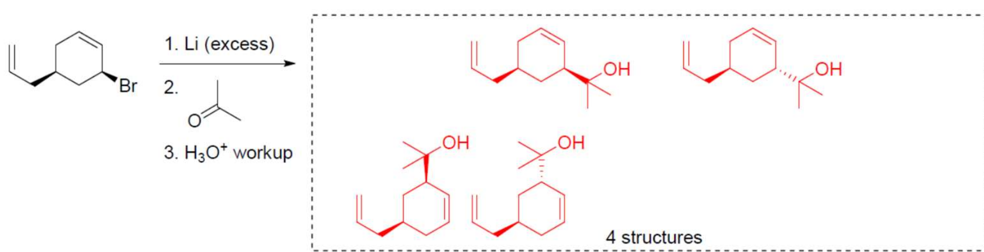
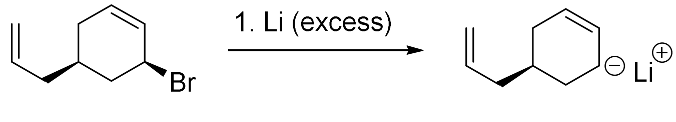
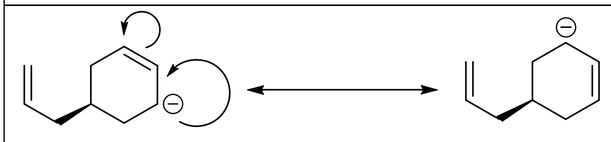
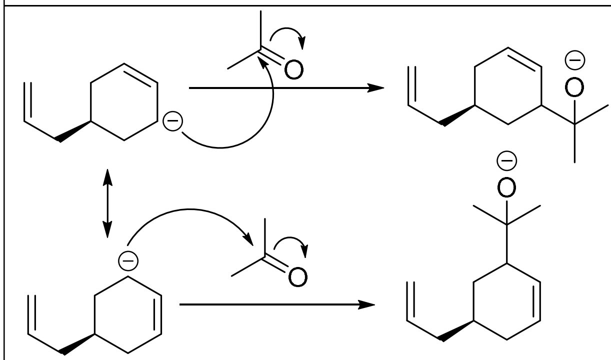
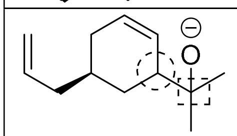
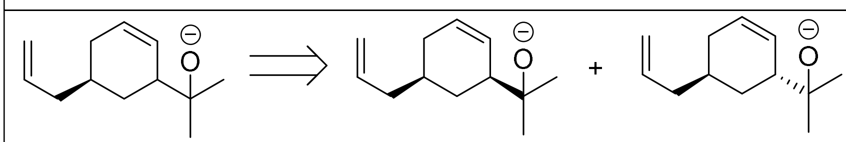
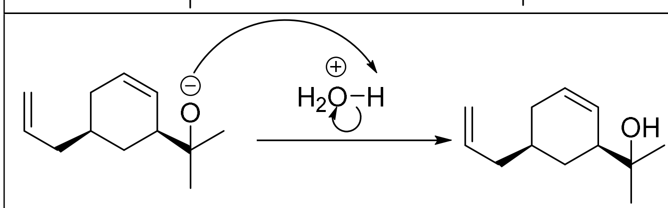


Table 8.3

Problem-Solving Process for Question 2: Exercise

Representation	Problem-Solving
	**Identify that Li (excess) will react with Br to form an organolithium reagent** - This will form a carbanion
	**Recognize that this is an allylic system so resonance is present.** - This means two reactive sites
	**Recognize that either position can act as a nucleophile and attach the ketone, resulting in two sets of products**
	**Identify the stereocenter (circled)** - Some students focused on the squared site which is not a stereocenter
	**Recognize that the stereocenter means that front and backside attacks will result in different products.**
	**Recognize that the H_3O^+ workup will convert the alkoxide into an alcohol.**
	Possible Distractors - Non-stereocenter formed by the reaction (squared) - Some students thought the wedged side-chain might be modified

Methods: Data Analysis

Video recordings and students' written work were collected from each interview. Video recordings focused on students' hands in order to capture their written work and gesturing while maintaining a higher level of anonymity. Interviews were first paraphrased which led to the choice to focus on questions 1 and 2. After initial paraphrasing student Interviews were then transcribed verbatim. These transcripts included a description of students' actions as they worked as well as their verbalizations. Preliminary thematic coding was then conducted on the interviews with a focus on what students did as they solved the problems (ie. identified reactive sites) and how they chose their next steps (ie. the reaction looked familiar). This round of thematic coding identified three key phases in student problem-solving which were in turn associated with a variety of reasoning strategies. These three phases consisted of decision points, brainstorming, and tension-points, and will be described in more detail later in the results section.

At this point activity logs were developed for the transcript of each problem. These activity logs broke down student think-aloud interviews into a series of brainstorming moments, decision points, and tension points (See Appendix F for an example). The activity logs included a description of each moment, the outcome, the reasoning strategies used by the student, a representative quote, and description of whether the moment put students on the right or wrong track to solve the problem. Being on the right or wrong track does not necessarily mean that the student showed understanding or even that their choice was reasonable, simply that it would have helped them get the correct answer. Six main categories of reasoning strategies were finalized. These strategies were not developed with a framework in mind though several of them did resemble strategies previously identified in the literature. The six main categories identified were: Recognition, Similarity, Undefined, Prediction, Chemical, and Unknown (See the Results section for more information).

Data from the activity logs were then tabulated in various ways. First tables were made for the number of brainstorming, decision, and tension points utilized by participants on each problem (See Table 8.4). Tables were also made on the decisions that students made for each question and the reasoning strategies that they used (See Appendix G). Tables were also made for the reasoning strategies used across all decision points and then those used at key decision points (See Tables 8.5 and 8.6). Finally a tables were made for the reasoning strategies applied at tension point, the outcome of the tension point, and whether this outcome lead participants in the right or wrong direction (See Table 8.7).

Results: Problem Solving Phases

Three phases of problem-solving were identified and refined through the initial thematic coding phase and the subsequent activity log phase.

Decision Points describe moments in which an individual makes a choice in their problem-solving. It can be between multiple options or simply a situation where a student selects their next step; decisions are often not contentious. These decisions are made using a variety of reasoning strategies and student will often use more than one strategy in combination. An example of a decision point can be seen when Keira decides that the carbonyl will reform and cause sulfur to leave because she was familiar with that pattern of reactivity.

Keira: I would probably use these electrons to make a double bond and kick that S over here just cause like looks familiar to something we might do.

In another examples William realizes that there is resonance in the molecule and so there are two reactive sites.

William: And you also notice there's a charge here so you can—you can do resonance around and then the charge will move over here. It can attack in a separate area.

Brainstorming is used to describe moments in a which a student remembers or considers an idea but no explicit decision is made about it. For example, in one of the interviews Nancy saw a portion of a molecule that reminded her of an amino acid but didn't think the information was relevant.

Nancy: This part looks like a—an amine acid because you got the N-terminus and then you got the C-terminus, although a part of it looks like a BOC group. I don't think this is useful to the predict the product part, but it's just stuff I'm noticing.

In another example, in one of the interviews Andrea identified and labeled a functional group. While this observation may have been applied to later decision the, act of identifying and labeling it was not a decision.

Andrea: Um so for this problem first I see a BOC group so I'm gonna go ahead and label that.

The identification of a brainstorming phase is in line with much of the problem-solving literature which has often identified a similar preparation phase(Bowen & Bodner, 1991; Larkin, Mcdermott, Simon, & Simon, 1980). Students tend to use this phase to define the problem space by describing aspects of the question that are called to mind. While brainstorming was often associated with various reasoning strategies, even when a final decision was not made, sometimes students brainstorming was purely **observational.** For example, commenting that part of a molecules was intimidating.

Tension points were moments in which students find that there is a conflict between different reasoning strategies. In these situations' students find that different reasoning strategies point toward different outcomes. Students in these moments must either try to resolve the two reasoning strategies or decide which reasoning strategy is more convincing. The resolution of tension points can provide more information about the reasoning strategies that students value and are most comfortable using. For example, Neil initially thinks that the amine nitrogen will react (instead of the amide nitrogen) but realizes that it doesn't allow him to make the products that he thinks he should make. In order to decide which position to react with he must first resolve the tension between two reasoning strategies

Neil: Um and in doing so (8.0) I have reason to believe that I am wrong. Um (2.0) I say that just because of the fact that if I'm doing everything over on this far side, it's a little bit harder to—to get this carbon dioxide to form and this alkene to—to become a thing.

In another example Amanda feels like the formation of an iminium species feels appropriate but hydroxide is a bad leaving group. In order to resolve this tension point she must decide whether the leaving group strength of hydroxide is more important or her familiarity with forming iminium species.

Amanda: I could see the loan pairs in nitrogen coming down and doing a double bond but OH is like not a good leaving group and that probably wouldn't be preferred.

Results: Reasoning Strategies

Six main reasoning strategies were also identified. One of these categories, **Unknown**, simply describes a situation where I was unable to determine why a participant made the decision they did. Also, while this section will describe each strategy separately students would often use multiple strategies simultaneously to make a decision. For example, a student might recognize part of a molecule and then use that to make predictions. A student might also initially identify a reaction because they remembered it from a homework question but then support that initial observation with a deeper conceptual understanding of the reasons why that reaction would be favorable. This would then make them more certain in their decision.

Recognition reasoning describes situations in which students recognize an aspect of the question such as a pattern of reactivity, a functional group/molecule, the property of a functional group/molecule, or a reactive site. Recognition reasoning refers to situations specifically where student recognizes an aspect of the question but this recognition is not tied to their understanding of the underlying chemical reasons or their memory of previous similar examples. Students in this situation seem to have abstracted the information and it's not tied to examples or reliant on a understanding of the chemistry. It is possible that students do understand the underlying chemical concepts but simply don't mention them. This however suggests that their knowledge of the chemistry is not central to their decisions.

For example, Patricia immediately recognized that a negative charge would form at the Br upon reacting with Li(excess) in Question 2 but does not indicate that she understands the reasons why this reaction would occur simply that it would.

Patricia: Okay um so, the first step is—becomes negative right here [points to Br]

Some participants seem to invoke this reasoning strategy deliberately. For example, Rebecca describes how she learned through her studying that it is important to recognize the acyl group and leaving group.

Rebecca: Oh one of the main things that I saw in my like studying was, I guess it didn't click with me until really into like my studying, that the acyl group is like super important to identify um and like what I would—what my leaving group would be.

Similarity reasoning describes situations in which students recognize a component of the question because of its resemblance to previous examples they've seen before in the course. Similarity is characterized by "this reminds me of" or "this happens a lot" statements. Similarity reasoning closely resembles case-based reasoning as described in Kraft et al in which students solve questions by comparing them to a previous case (Kraft et al., 2010). For example, Keira indicates that she believes the amide would want to attack something because it is a reaction she has seen frequently and not based on any reasoning about nitrogen's reactivity.

Keira: I would probably think that the lone pairs on the N would want to attack something just because that happens a lot.

Or in another example, when Nancy begins question 1 she immediately feels that the thioanhydride group reminds her of SO_2.

Nancy: So this part [referring to the thioanhydride functional group] reminds me of this molecule [draws O=S=O] that's like created and drives reactions.

Undefined reasoning describes a situation in which a student makes a decision because "it feels right" or "doesn't feel right." It also includes situations where students make a best guess. Students in this situation are not specifically calling upon specific memories or cases but a vague sense of familiarity. For example, when Keira adds a hydrogen to a sulfur leaving group simply because she felt like it should do that without considering any reactions that would lead to that addition to occur.

Keira: And then honestly, I would not know what do from here, and I would probably just add an H because I know you probably don't leave a minus.

Predictive reasoning describes situations in which students try to make predictions on what might occur based on assumptions of what the final product should be. Predictive reasoning is equivalent to the means-end analysis previously described in the literature (Bhattacharyya & Bodner, 2005; DeCocq & Bhattacharyya, 2018; Larkin et al., 1980). Students will sometimes use hints, in the form of information given to them by the instructor in the product box, to help them in problem-solving and in making decisions. For example, on Question 2 Katie immediately realized that the hint 4 structures meant there would probably be stereocenters in the products.

Katie: Okay so the four structures, so that means stereocenters.

Students will also sometimes utilize their familiarity with the starting materials and reagents given in the problem to make predictions. For example, in question 1, Neil recognized the BOC group and decided that the reaction must result in the removal of the BOC group.

Neil: Um I—I remember—I remember seeing this side chain [circles BOC protected amide] um I think it was one of the lectures, uh and I specifically remember uh having it break up such that you form CO_2 [draws O=C=O] and [draws an alkene] and (1.0) alkene. Yean and uh—and an alkene. Um (2.0) and so I think, knowing that, the—the problem becomes much easier.

Chemical reasoning in contrast in when a student considers underlying chemical principles when brainstorming or making a decision. Students in these questions make their decision because they understand why the step occurs based on underlying chemical concepts. They have not abstracted the information and it's not based on how similar the question to another example. For example, in Question 1 when choosing whether the amine or the amide nitrogen would act as a nucleophile Andrea considered their relative nucleophilicity, though she mistakenly spoke about this in terms of basicity.

Andrea: I know that an amide is already, base, is—is already very conjugated stabilized so I'm not gonna use that either, but this amine looks like a good base do I'll go ahead and use that.

In order to be classified as chemical reasoning students did not need to be thinking about accurate chemical concepts, and student could still be using chemical reasoning but have misconceptions. For example, Brianna considers what charges would form following a reaction but doesn't understand how charges get formed. Her final decision that taking electrons away forms a negative charge is actually incorrect but still demonstrates that she was considering the movement of electrons when making her decision

Brianna: Um but just like the arrow pushing still confused me like I guess that would be a plus charge [draws a plus charge on the amide N]. Is it a minus? I don't know that still confused me. Like if you're pushing a bond into something I think that makes a plus charge but if you were to take the electrons away it is—would make a minus, or I have it backwards.

There is a wide range of chemical concepts that a student could potentially consider but the following areas were identified in student interviews: Electrophile/Nucleophile Properties, Acid/Base Properties, Structural Properties, Mechanistic properties, and Other properties. **Electrophile/Nucleophile** Properties involves situations where students discussed whether or not a reaction between an electrophile and a nucleophile was favorable. This could for example include thinking about leaving group strength. **Acid/Base** Properties usually was situations in which a student discussed whether a reaction between an acid and a base was favorable. **Structural** properties involved students making decisions based on structural properties of a molecule, this could involve students identifying symmetry, stereocenter, or resonance. **Mechanistic** properties involved situations where students discussed whether there was a reasoning mechanism for a transformation. This was often discussed by students in terms of whether there was a "driving force" available for a reaction. Finally, **Other** included any other types of chemical reasoning that appeared in the interviews. For example, there were a couple cases where students discussed charge balancing or periodic trends.

Data Analysis: Navigating the Problem Space

Overall students are much more able to identify and focus in on relevant ideas in the exercise-type question than in the problem-type question. This can be seen in the lower number of decisions, brainstorming moments, and tension-points in Question 2. Additionally, a much higher number of different types of decisions were identified in Question 1 than in Question 2. There was also a wider range of ideas considered in tension points and brainstorming moments. Overall this seems to indicate that there is a much larger problem space in Question 1 than for Question 2, and students therefore have to consider and sort through a wider range of ideas.

Table 8.4

Number of Brainstorming, Decision Point, and Tension Phases

	Question 1 - Problem			**Question 2 - Exercise**		
	Brainstorming	Decision	Tension	Brainstorming	Decision	Tension
Neil	0	6	2	0	5	0
Andrea	1	10	2	1	4	1
Katie	4	3	1	0	6	0
Patricia	0	6	0	0	6	0
Nancy	6	7	3	0	6	0
Rebecca	2	8	2	1	8	3
Brianna	5	6	1	7	6	3
Keira	0	6	1	1	5	0
Andrew	4	3	1	1	6	2
Tessa	3	5	1	1	8	0
William	1	7	2	2	7	3
Sophia	2	8	1	3	9	0
Olivia	8	8	3	1	5	1
Samantha	2	9	2	2	5	0
Amanda	6	9	2	0	6	2
Total	44	101	24	20	92	15
Average	2.9	6.7	1.6	1.3	6.1	1.0

Participants on average made 6.7 decisions in question 1 and 6.1 decisions in question 2 (See Table 8.4). This difference is fairly mild but a much larger difference can be seen in the number of brainstorming moments and tension points that arose between the two questions. Students had notably fewer brainstorming moments and tension points arise in the exercise than in the problem. Students in question 1 had, on average, 2.9 brainstorming moments, while in question 2 they only had 1.3. Students in question 1 on average had 1.6 tension points and 1.0 on question 2. In addition, only seven students had tension points arise in their problem-solving for the exercise-type question, while 15 students had tension points arise in their problem-solving for the problem-type question (Table 8.4). All of this suggests that students were able to get to an answer in fewer steps—after considering fewer ideas—in the exercise than in the problem.

This is further supported when we look at the nature of the decisions that students made on each of these questions. Notable differences were seen in the range of ideas that students use when making decisions in question 1 and question 2. Analysis of students' decisions indicated that in the problem-type question students made a total of 53 different types of decisions (See Appendix G). This list included common decisions such as: "The amide N reacts the with carbonyl and forms a tetrahedral intermediate" and "Pyridine is a base". It also included more unusual decisions such as "Sulfur in the ring is protonated". Taking a break or giving up were also included as decisions. Students also considered a wide range of ideas in their brainstorming and tension-points that never made it into their decision making (n = 38). The exercise-type condition, in contrast, resulted in only 34 different types of decisions, and there were far fewer ideas considered in their brainstorming and tension-points (n = 26). The large difference in the range of decisions made between the problem and the exercise suggests that the problem space for the problem is much large than it is for the exercise. Students seem much better able to constrain the problem space for the exercise and focus in on relevant ideas. In contrast, students must sort through a wider range of ideas in the problem-space resulting in more variation in the decisions that students make.

This broad problem space meant that participants sometimes struggled to identify the relevant information and correct approach to solving question 1. This broad problem space sometimes led students to focus on incorrect or irrelevant information and make it central to their problem-solving approach. In this situation students often used similarity reasoning to identify a part of the question that was similar to an example they had seen previously in the course and then made predictions about how the rest of the problem should progress based on that piece of information. This combination of similarity and predictive reasoning often led students to become focused on irrelevant or incorrect information even when their decisions resulted in tension points. It would also lead them to ignore proposed ideas that would have been more correct. This approach was seen in both high-performing and low-performing students. The following section will discuss three cases in which students engaged in this type of problem-solving behavior.

Navigating the Problem Space: Neil

Neil is a high performing student who described enjoying the Organic Chemistry 2. He had received an A+ in Organic Chemistry 1 and his final grade in Organic Chemistry 2 was an A. When first approaching question 1 Neil described not being quite sure how to start. However, he quickly recognized the BOC sidechain from an example he'd seen in the course where he had to break it up to form CO_2 and an alkene.

Neil: Yeah this one—this one was interesting, because I think when I—when I first looked at this uh I was not quite sure where to start. Um I—I remember seeing this chain [circle BOC group' um from I think it was one of the lectures uh and I specifically remember uh having it break up and such that you form CO_2 [draws O=C=O] and [draws an alkene] an (1.0) alkene. Yeah. And uh—and an alkene. Um (2.0) and so I think, knowing that, the—the problem becomes much easier.

While students did see examples where a BOC protecting group was removed in a problem set, this reaction required the use of a strong acid, like HCl or trifluroacetic acid, not a weak base like pyridine. This means that under the conditions provided this would not have been a feasible reaction. However, Neil continued to make decisions based on this assumption. While he originally correctly identified that the amine would react with the carbonyl, he then ran into a tension point where he realizes that it was far away from the BOC group and therefore the amine reacting with the carbonyl would not help him in his goal of removing the BOC protecting group.

Neil: You can just attack one of the carbonyls [draws arrows from amine nitrogen to carbonyl] to form something along those lines [humming]. Eh, yes yes. Um and in doing so (8.0) I have reason to believe that I am wrong. Um (2.0) I say that just because of the fact that if I'm doing everything over on this far side [points to amine] it's a bit harder to—to get his carbon dioxide to form [points to amide bond] and then alkene [points to tertbutyl group] to—to become a thing.

Neil therefore made a mistake in solving this question, based entirely on his goal of removing the BOC protecting group and creating CO_2 and an alkene. He, however, did not consider whether the amine or the amide nitrogen would be more a nucleophilic site (better able to react to with a carbonyl).

This focus on removing the BOC protecting group even lef him to ignore evidence that the reaction might not be feasible. Upon reaching the point in the question where he needed to remove the BOC protecting group, Neil found that he was unable to determine a driving force for the reaction. This resulted in a tension point in which he struggled between his certainty that the goal of the question is to remove the protecting group and his inability to find a mechanism for how the reaction would occur. Neil ended up deciding that since he was not required to show a mechanism for the question it was not important that he know why the reaction occur, simply that it does occur.

Neil: Um looking at this (6.0) my thoughts are (3.0) that I should be (2.0) wondering about this. Um I, at a glance, don't really see any (2.0) intermediate or like driving force. Because I understand what needs to happen is this—this pair of electrons needs to go over [draws arrow from C-O bond to C-O bond] (help) from the carbon dioxide [draws arrow from C-N bond to N]. This needs to happen. Um that needs to give its electrons back to nitrogen. Um (3.0) but I don't necessarily see any particular driving force for that. Um (3.0) and so (3.0) with that in mind, even though I know what must happen um I question whether I know how it happens. Um but given that this is just predict the product um (1.0) the—the mechanism itself is not necessarily the most important thing um so much as that you get the—the product out of it. So, I'm just going to magically say that that happens because science.

Neil's initial use of similarity reasoning in connecting the BOC protecting group in question 1 became central to the rest of his problem-solving and he made many key decisions based on this goal. This resulted in several tension points where other reasoning strategies came in contrast with his prediction. However, he chose to either modify his decisions or ignore potential issues in order to stay in line with his prediction. This approach led him to get the question wrong because his overall assumption was incorrect.

Navigating the Problem Space: Nancy

A similar problem-solving approach was seen when observing Nancy. Nancy started the semester as an average student having received a B in Organic Chemistry 1 and ended Organic Chemistry 2 with an A-. Nancy began the problem by defining the problem space, brainstorming about the various aspects of the problem that seemed familiar to her and considering the hints that had been given the problem. After this process however she decided that the oxygen in amide carbonyl would react with the sulfur in the thioanhydride species because this was something that she had seen in previous questions.

Nancy: So, I took the O and I attached it to the sulfur because I knew that was a thing that happened with other problems that I'd seen.

While students had previously seen examples where oxygens and sulfurs reacted with each other, they were relatively rare and were only seen in reactions with thionyl chloride, a reagent that is used to hydroxide groups with chlorides. Notably in reaction between oxygen and thionyl chloride the oxygen's attack on sulfur results in a chloride group leaving. The reaction Nancy proposed between the carbonyl oxygen and thioanhydride sulfur, in contrast, did not include a leaving group. Nancy's choice seemed overall motivated by similarity reasoning but was hindered by the fact that she had an incomplete memory of the reaction resulting in her basing most of her reasoning in this problem off of the ideas that, "Sulfur reacts with oxygens."

Shortly after proposing this reaction Nancy took a break from the problem, but upon returning on it reiterated her belief that the sulfur and the carbonyl oxygen would react. She then continued to make decision based on the idea that goal of the problem was to have the oxygen on the carbonyl react with the sulfur. Her next significant decision was that pyridine would deprotonate the amide nitrogen in order to provide a driving force for the reaction. This resulted in her choosing to deprotonate the least acidic of the two nitrogens because it helped her to advance the react that she believed would occur.

Nancy: Then I know that there has to be like a driving force behind why this goes here [draws arrow from C=O to the oxygen] so I would probably take this H using the base [draws pyridine reacting with amide H]. I guess I don't know that there needs to be a driving, like maybe there doesn't need to be a driving force but in the practice one's that I did and stuff like that this just doesn't normally happen on its own so I would probably take this hydrogen and then move this double bond up her [draws arrow from N-H bond to N-C bond].

Shortly after this moment though Nancy reached a tension point where she realized that the amine nitrogen could also be deprotonated. She considered both options but decided that deprotonating the amine didn't make sense because there would be a negative charge on the nitrogen instead of the more electrophilic oxygen. This decision however ignored that the amide nitrogen would be less acidic because of conjugation with the carbonyl.

Nancy: I would be stuck between either this one [circles product] or deprotonating here first [points to amine in starting material]. So then I would take hydrogen from here [draws pyridine deprotonating the amine N] and then I would have NH minus but then that doesn't make sense because nitrogen doesn't like the negative charge at all.

This tension point is unable to help direct Nancy in a better direction because she didn't completely understand the factors that lead to differences in acidity. Follow-up questions revealed that even in a situation where she would deprotonate the amine first she still considered a reaction with sulfur to be the overall goal.

Interviewer: So if you that one—um so you said you'd be torn between then, if you had that where do you think it would want—where would—what would you want to do next with it?

Nancy: If I had this one, I would want to get rid of this negative charge as quickly as possible, so I would probably attach—see I would attach the nitrogen to the sulfur [draws an N-S bond], but then it would feel wrong because I've only ever attached sulfurs to oxygens, um so then I would probably scrap this [crosses out N-S bond] and go with this one [points to product]

Nancy's initial incorrect decision, motivated by similarity reasoning, that the oxygen would react with sulfur remained pervasive throughout her problem solving. It motivated her to believe that the least acidic site would be deprotonated. Then when a tension point arose due to her noticing the amine site, she lacked understanding of the chemical concepts underlying acidity. These misconceptions strengthen her belief that the amide would be deprotonated. Even when prompted to consider a hypothetical scenario where the amine is deprotonated, she still proposed a reaction with the sulfur and did not consider a reaction with the carbonyl. Since reactions between sulfur and nitrogen had never been seen in the class she returned to her original answer.

Navigating the Problem Space: Olivia

Olivia struggled in Organic Chemistry 2 and described finding the class frustrating. She had received a B- in Organic Chemistry 1 and ended up receiving a C in Organic Chemistry 2. She, like Neil and Nancy, struggled to identify relevant concepts in a broad problem space, and chose in the end to focus on a partially remembered idea from class. Olivia, like Nancy, began the question by attempting to define the problem space, focusing particularly on the functional groups in the starting material and the role of pyridine. After noting that sulfur's a "natural withdrawing group" she decides that she would protonate the sulfur because it reminded her of a similar reaction.

Olivia:	Oh okay so we can protonate the sulfur [points to sulfur]. I think um I forget why, we can and pffh yeah
Interviewer:	What are you thinking?
Olivia:	Using what?
Interviewer:	Oh I was just gonna say what are you thinking?
Olivia:	Um cause I remember one of the reaction schemes you protonate the—you protonate the oxygen that's in the ring and then that breaks the ring, but I forget the rest of it.

Olivia's choice to protonate the sulfur was based on her observation that the question resembled previous examples where oxygen is protonated and this resulted in the ring opening. Her belief that oxygen and sulfur would be behave similarly was well founded and her answers to later reflection questions indicated that the she has a good understanding of periodic trends.

Interviewer:	Uh what did you do that helped you prepare for a problem like this?
Olivia:	Um again, like I said before I recognize that's this um—sulfur and oxygen are in the same like column on the periodic table so they behave similarly. Um I mean there's a mismatch of like p orbital right now um but that's why I treated it as if it had been an oxygen in the ring which is where I applied the mechanism that I remembered but then forgot right now.

However, it was unusual that she chooses to focus on this particular reaction since it would occur under acidic conditions and question 2 is occurring under basic conditions. Olivia in her brainstorming had identified the presence of a base, pyridine, in the reaction, but did not to connect that these conditions meant that protonating sulfur would be unlikely. Olivia at this point seemed to have decided that opening the ring is the point of this problem and that this would require protonation even under basic conditions. She also misremembered the reaction and chose to protonate the sulfur instead of the carbonyl. Later she realized that the carbonyl could be protonated which resulted in a tension point. However, she incorrectly reasoned that this would not be helpful and then got confused in her explanation.

Olivia: Yeah the other option would be to protonate one of the carbonyl oxygens, but I don't know if that would help me. I mean the reason that you deprotonate this one [points to pyridine] is because this one has the double bond here [points to carbonyl] which—and this one has the two loan pairs [draws loan pairs on sulfur] so that means it would be—pretty that it means it acts—I don't know what means I'm sorry

Olivia finally decided that she would protonate the sulfur group and that the proton would come from the amine position. While she was correct that the amine hydrogen is the most acidic, the reaction as described is not possible since the sulfur in the thioanhydride would not be a strong base. Olivia also did not indicate any rationale for her choice to deprotonate the amine, so it is unknown if this was due to it being more acidic.

Olivia: Okay so if I'm protonating this one [draws H on S] then that means that this will turn into NH [redraws amine species with one fewer H]

Olivia then struggled to determine the next step. She identified that the deprotonated amine was a strong base as well and then found that she was unable to move arrows in a way that she found convincing. She chose to take a break from the problem and then returned to it later. Upon returning to the problem she again observed that the deprotonated amine is a strong base and that that could be significant. This resulted in a potentially productive tension point in which Olivia considered that the amine could react with the carbonyl.

Olivia: Uh okay so actually this is going to act like a strong base [circles amine] this because we have our um (3.0) amide—not amide—imine. Is that imine? Yeah I mean it has—not that's not what it is. Anyways it's a base um so it's going to act like base (4.0). What does that mean? Um I mean it means that it could attack the carbonyl directly but like I don't know if it's going to do that because it's a ring um (2.0). I could try it. That's going to push that out there [gestures toward the question]. I don't know if that helps, then we just have a negative (4.0) with no hydrogens anywhere which doesn't help me, so I think I'm going to go with my original which is the H [draws H on S].

Unfortunately, while Olivia realized that deprotonated amine could react with the carbonyl, she was unable to completely drop her original idea. She became concerned that the reaction between the amine and the carbonyl might not occur within a ring and that it might be unhelpful in opening the ring. When she described the resulting alkoxide that would form she was concerned by both the negative charge and the lack of hydrogens. She seemed to hold on to her initial similarity reasoning and when the reaction between the amine and the carbonyl did not resemble it dropped it completely. In the end she was unable identify a mechanism to open the ring and ended up giving up the problem.

Navigating the Problem Space: Conclusions

While Neil, Nancy, and Olivia all had different levels of comfort with the material they all struggled with the same issue. The broad problem space included many ideas that they could consider and they picked one that seemed similar to examples that they had seen previously. They then based all their subsequent decisions on that idea. Their memory of these examples was usually incomplete but they remained focused on them even when tension points arose that indicated that it was not a successful strategy or that alternative strategies should be considered. We can see in both the cases of Nancy and Olivia that even when considering alternative ideas, they were unable to completely give up their initial idea and it led them to drop potentially promising pathways because they insufficiently resemble the examples they had in mind.

In the exercise-type question however students were able to narrow down the problem space more easily. Most students were able to recognize that this question was the reaction of an alkyl lithium with a ketone and did not propose alternative routes. Questions of this type were very familiar to them and they had seen very similar questions in lecture and in their homework. They had even seen simpler versions of them in Organic Chemistry 1. Students, in general, were able to recognize the three main reactions and only struggled in identifying resonance and the creation of stereocenters.

Data Analysis: Key Decision Points

Differences, however, still arose between the problem-type question and exercise-type question even when students were able to focus in on relevant information and avoided more unusual pathways. Even within these more reasonable pathways, students still needed to make key decisions that influenced whether they ended up with correct or incorrect answers. At these decision points students could use a variety of different reasoning strategies but results indicate that the problem-type question required the use of different strategies than the exercise-type question.

Table 8.5

Use of Reasoning Strategies during decision points

	Question 1 - Problem		**Question 2 - Exercise**	
	Total (across population)	Average (per person)	Total (across population	Average (per person)
Predictive	25	1.7	21	1.4
Recognition	26	1.7	37	2.5
Similarity	18	1.2	13	0.9
Undefined	11	0.7	6	0.4
Chemical	30	2.0	22	1.5
Unknown	7	0.5	1	0.1

A broad overview of student reasoning strategies when making decisions reveals differences in the reasoning strategies used in the two questions (See Table 8.5). When making decisions in question 1, students most frequently used chemical reasoning although predictive and recognition reasoning were also quite common. However, in question 2, recognition reasoning was much more common than any other strategy. This seems to indicate that students are better able to recognize key features of the problem when it's an exercise but in problems are unable to and therefore use other strategies more equally. Analysis of question 2 supported this and I found that, in general, students were able to recognize the three main reactions – and only struggled in identifying resonance and the creation of stereocenters. In question 1, as we saw in the previous section students struggled more to identify relevant reactions and used a wider variety of strategies.

More in-depth analysis of these decision points shows that students not only use different strategies in the two questions but that different strategies are most effective in each question. While students make a variety of different decisions, in each question I was able to identify one key decision point that influenced whether or not students were able to get the correct answer to the question. These key decision points were characterized by being common decision points (one most of the participants would have had in their problem solving) with multiple outcomes and a significant impact on the outcome of the question. They are also moments where students didn't always make the correct decision and where participants had used a variety of different reasoning strategies. The following sections will discuss analyses of the two key decision points and how students use of reasoning strategies influenced their performance on each question.

Table 8.6

*Use of Reasoning Strategies at Key Decision points and outcomes of decision**

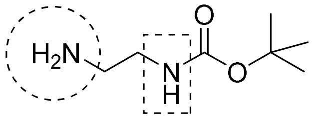
	Question 1 - Problem		**Question 2 - Exercise**	
	Correct	Incorrect	Correct	Incorrect
Predictive	0	1	4	1
Predictive/Chemical	0	1	3	0
Recognition	1	0	0	0
Similarity	1	2	0	0
Undefined	1	2	0	0
Chemical	4	0	3	1
Total	7	6	12	2

*Students who reasoning strategy was unknown or who did not make this decision were not included in this table.

Key Decision Points: Problem-type Question

In the problem-type question I identified the key decision point to be deciding which nitrogen would react (See Figure 8.1). Participants had to choose whether the amine position (circled) would be the reactive site or the amide position (squared). Students also had to decide whether the position would react either with the pyridine or the thioanhydride first, but since the order did not impact the final product, this was not considered a key decision point. In order to get the correct answer students needed to identify that the most reactive nitrogen was the amine. The amide nitrogen, due to the presence of the BOC protecting group, was stabilized by conjugation (the sharing of electrons) and therefore unreactive

Figure 8.3 Possible reactive sites



While students had been introduced to the idea that amides were less nucleophilic and less acidic, and more specifically that BOC protecting groups would make an amine less nucleophilic, they rarely came across questions, where they had to make this decision. It was therefore unlikely that they would have immediately recognized the more reactive site or could have called upon a specific example case. This meant that in order to get this answer correct students needed to understand chemically why the amide nitrogen was less reactive than the amine nitrogen. This is reflected in student reasoning at this decision point. The four students who utilized chemical reasoning to make this decision all got the answer correct (See Table 8.6). Interestingly those who did not engage in chemical reasoning tended to select the wrong nitrogen, particularly when they utilized predictive reasoning, similarity, or undefined reasoning. This seems to emphasize that in this question students weren't able to develop heuristics or instincts for how this step should proceed and when they tried to do so were unsuccessful.

We can see in the case of Andrea how chemical reasoning allows participants to accurately select the reactive site. Andrea begins her problem solving by considering the reactivity of the various functional groups in the amide species. She observed that the BOC group will not be reactive, neither will the amide nitrogen because it's conjugated, and then identifies the amine as the reactive site due to it being a good base. While her terminology is incorrect, technically in this reaction the amine is a nucleophile, the characteristics that make the amine a good base are also what make it a good nucleophile.

Andrea: So I know that the BOC group, because it's a protecting group, is not going to involve itself in any chemistry so I'm not gonna use it. I know that an amide is already [points to amide], base, is—is already very conjugated stabilized so I'm not going to use that either. But this amine [points to amine] looks like a good base so I'll go ahead and use that.

Another successful strategy can be seen in the way that Sophia approached this decision point. While Sophia does not consider how conjugation would make the amide nitrogen less nucleophilic, she remembers that the BOC group is a protecting group and will therefore make the amine nitrogen less reactive. This allows her to correctly identify the most reactive site.

Sophia: Um well for this one, this part of it over here [points to BOC protecting amide group] um is like that protecting thingy so I figured this nitrogen and the hydrogen [points to amide N] wouldn't react because it was being protected um and since that was the thing in this problem that I thought was going to react, like one of the nitrogens, I figured it was going to be this one [underlines amine] not that one [points to amide N].

Some students were also able to identify the most reactive site without utilizing chemical reasoning. In some of these situations the participant appears to not notice that the amide might be a reactive site. For example, Katie simply points out the amine is the reactive site but did not explain why or why she did not consider the amide nitrogen. In other situations, students seemed to arrive at the right nitrogen somewhat by luck. Some, like Samantha, simply guessed and happened upon the correct position.

Samantha: And I'm like well I need to pull hydrogens off somewhere and the places that I obviously see are like [circles the two N positions] hydrogens sticking here. I would be like there's a possibly that I would pull one from one of these carbons but in my head it's like less likely and so just because it comes down to a very arbitrary decision where I'd be like okay I'm just gonna take this guy [points to hydrogen on amine] and make sure that he ends up over here [points to protonated pyridine]

However, strategies like similarity and undefined reasoning tended to result in students selecting the amide position (See Table 8.6). This suggest that the correct choice of the amine is likely chance and not evidence of a well-developed intuition or recognition of reactivity patterns. Several students utilized undefined or similarity reasoning to explain why they felt that the amide was the correct choice. For example, when asked why she chose the amide site Keira describes that she did so because it only had one hydrogen attached to it and that made the option feel right to her. She then expands to say that if the product of the product "looked weird" she would then consider the amine site.

Interviewer: Okay cool um how did you decide which nitrogen to attack with?
Keira: Oh mm that's a good question (3.0) [sigh] (5.0) I don't really know. I literally just chose it because t had one hydrogen attached which really doesn't have to do with anything. That's just like—that's what I—that's the reasoning why I chose it.
Interview: That's fair, cool.
Keira: But then I don't know if I was unhappy with this or I thought it looked weird then I would try like [points to amine] attacking it with another one and seeing if I thought it looked better. But that's not with like much reasoning. That's just like intuition or like mm I've seen this or I haven't seen this.

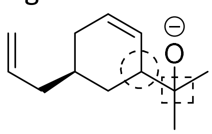
Tessa also chose the amide position using ill-defined similarity reasoning. Interestingly, Tessa held on to her choice even when she began considering chemical reasoning for her choice. She quickly dismisses chemical reasoning however and reiterates that she remembers that you use the amide position.

Interviewer:	Okay um, so how did you decided which um like nitrogen to attack with?
Tessa:	I remember that—well this is the amide one [circles amide group]. So usually you do that. Um let's see (2.0) and I guess (1.0) hmm (2.0). I'm not sure if like a resonance argument might apply to it, which would make the—that nitrogen. Well then it mm be more mm [laugh]. I just remember that you do the—the amide one.

Overall it seems that students needed to understand the underlying chemical concepts in order to correctly identify the amine as the reactive site. Students who utilized other strategies, particularly similarity, undefined, and predictive reasoning, were not able to consistently identify the reactive site. Many seemed to find reactions at an amide nitrogen more familiar, even though it was not a reactive site. It is also interesting that some students would stand by their belief that the amide site was the more reactive site even being willing to dismiss chemical reasoning that might cause them to question their decision.

Key Decision Point: Exercise-type Question

Figure 8.2 Potential stereocenters in intermediate



In the exercise-type question the key decision point was identifying the creation of a stereocenter. The creation of a stereocenter meant that attacks from the top and the bottom would result in different products (indicated by wedged and dashed bonds). Most students were able to identify that stereochemistry was relevant and would result in the carbanion attacking from the top and bottom and forming both wedged and dashed products. The main sources of error were students who assumed that the alkene side-chain would change its conformation and those who were confused because the alkoxide carbon was not a stereocenter (squared) and missed the stereocenter that formed on the carbon on the ring was (circled) (See Figure 8.2). Students had previously seen many similar examples of questions where they needed to create alkyl lithium species and then react them with ketones. These types of questions had even been seen in Organic Chemistry 1. Students were also accustomed to seeing hints about the number of structures and had therefore had opportunities to develop heuristics for solving these types of questions.

This meant that unlike in question 1 students did not necessarily need chemical reasoning to consistently get this answer correct (See Table 8.6). Many students did use chemical reasoning and specifically identified the stereocenter. For example, while Rebecca is initially confused about the location of the stereocenter and even considers changing the orientation of the alkene side chain she eventually identifies stereocenter and is able to predict all the correct products.

> Rebecca: Oh and no wait, yes, this a stereoisomer [points to the carbanion in the first intermediate] so it can attack from front of back so that's gonna give me two products here.

However, many students only used predictive reasoning to make this decision, assuming that four structures mean stereocenters and drawing in the wedges and dashes without verifying that a stereocenter was actually formed. Katie for example begins solving question 2 by declaring, "So the four structures so that means stereocenters." Then at the end of the problem simply describes that there will be "different versions" of the products and draws the products from frontside and backside attacks. She never uses chemical reasoning to identify a stereocenter she is simply aware from previous examples that one would be present.

Another viable strategy was a combination of chemical reasoning and predictive reasoning. In these cases, students realized that a stereocenter would be present because of the predicted product, but then specifically seek it out. When asked about how she recognized the stereocenters Olivia describes it as a cyclical process where seeing four structures prompts her to look and see if there is a stereocenter present.

> Interviewer: What—what helped you recognize the two stereocenters?
>
> Olivia: Um the fact that there were four structures. It's kind of—it's kind of cyclical. Cause I'm like really—really what I did was asked myself the question. Okay so there are four structures so like I know that when they're asking for multiple structures sometimes that means that there's probably some stereochemistry going on, like you create a couple of different conformers and that leads to more products…It's just like (2.0) I ask myself the question like, "Hmm is there any stereochemistry going?" And originally I was like, "Oh my god that's not a stereocenter" but then I'm like, "But wait that's not the stereocenter."

While most students made the correct decision at this decision point there were two who did not get the correct answers. One student Samantha made a mistake in the reaction between the carbanion and the ketone resulting in her proposing an unusual structure with an extra carbon atom. While she did utilize predictive reasoning to identify stereochemistry her product was fairly unusual and therefore her choices at this decision point are difficult to compare to other responses. Andrea however used a combination of predictive and chemical reasoning to try to identify structural characteristics of the molecule that might result in multiple products. She however is distracted by the alkoxide carbon not being a stereocenter and is unable to identify the stereocenter at the carbanion site even though she considers many different possibilities.

Andrea: And I'm like where are my four different structures because this is not a stereocenter because I have uh two methyl groups attached to it [points to alkoxide carbon] so I'm like did I miss something really big…I don't think any E1 or E2 is happening, I don't think we're switching around any double bonds. And, yeah I'm just going to write down this answer because it's the only thing I can think of and I'm running out of time on this exam.

Several students also initially approached this decision point incorrectly and then in a subsequent tension point identified a better choice. These students, such as Rebecca and William, initially used predictive reasoning to propose that the additional products must come from changing the orientation of the alkene side-chain. However subsequent reflection made then realize that the alkene sidechain would be unchanged and that instead a stereocenter would be present at the site of the carbanion.

William: And that's just uh one of the product because you see uh there four structure [underlines 4 structures] so you have to make—to notice that. And one of the ways to do that is the stereoisomer around here. So I guess you can draw this [draws a product with a dashed alkene sidechain] that's another stru—possible structure…Actually let's see, I think no thing—thing—this stays her actually [converts dashed alkene sidechain into a wedge] and just this change [draws a wedged bond to the other sidechain].

In exercise-type questions students were able to use predictive reasoning with a high level of efficacy, and often used a combination of predictive and chemical reasoning where they using predictive reasoning based on the hint "4 structures" to propose that there was stereochemistry and then looked for stereocenters on the molecule. Overall this indicates that in an exercise-type question students' usage of predictive reasoning can be just as effective as chemical reasoning. While students use of predictive reasoning seemed to sometimes point toward changing the orientation of the alkene side-chain, students who used chemical reasoning sometimes risked missing the stereocenter altogether.

Key Decision Points: Conclusions

Overall these results indicated that for participants chemical reasoning was a very important strategy in solving question 1, particularly at the key decision point. In contrast chemical reasoning was less important in question 2. The most common reasoning strategy was recognition and the key decision point could be solved correctly by either predictive reasoning or chemical reasoning. This may suggest that problems may be more effective at measuring student understanding of chemical concepts because the similarity of exercises to previous examples allows them to use alternative reasoning strategies.

Results and Data Analysis: Tension Point Resolutions

In the previous section it was shown that exercises and problems differed in the extent to which students needed chemical reasoning to solve the problem. Exercises in many ways do not require the use of chemical reasoning, whereas chemical reason seems much more critical in problem-solving settings. Throughout this chapter I have also discussed how students have reacted to tension points that have arisen in their problem solving and their tendency to often fall back on similarity and predictive reasoning even when exposed to contrary chemical reasoning. In this section I will explore in more detail how students resolve tension points and in particular whether chemical reasoning is valued at tension points and whether it is an effective reasoning strategy.

Tension points are an interesting place to explore student decision-making and reasoning strategies because they are environments in which students varying reasoning strategies come in conflict and students have to select which one's they find more convincing. This provides information on which reasoning strategies are most effective but also which strategies students default to. This section will predominantly focus on tension points that included chemical reasoning as this was an area of interest in the previous section and is reflective of the deeper conceptual understanding that is emphasized in the chemistry education literature ((Nakhleh & Mitchell, 1993; Raker, Holme, & Murphy, 2013)

Twenty-two tension points were observed across participants for Question 1-Problem, of which 17 included tensions between chemical reasoning and at least one other type of reasoning (often predictive, similarity, or undefined). Fifteen tension points were observed across participants for Question 2-Exercise, of which 10 included tensions between chemical reasoning and one other type of reasoning (usually predictive). While almost all tension points were resolved, there were four recorded tension points in which participants were unable to find a resolution. In these situations, students often gave up and were unable to reach a solution. Only one of these unresolved tension points involved chemical reasoning.

In Question 1 participants chose chemical reasoning in 53% of relevant tension points (n = 9), while on Question 2 the only chose chemical reasoning in 50% of relevant tension points (N = 5). This overall suggests that chemical reasoning is not consistently viewed as the most convincing strategy and suggests that students do not trust in their ability to engage in chemical reasoning. This attitude can be seen in earlier cases where students chose to continue utilizes partially remembered example problems even when they clashed with chemical reasoning.

Neil as we saw previously at one point while trying to solve question one was trying to decide between predictive reasoning, his previous prediction that the BOC protecting group would leave, and chemical reasoning, his inability to identify a reasonable mechanism that would lead to this reaction occurring.

> Neil: Um looking at this (6.0) my thoughts are (3.0) that I should be (2.0) wondering about this. Um I at a glance I don't really see any (2.0) immediate or like driving force because I understand that what needs to happen is this-- this pair of

electrons needs to go over [draws arrow between two bonds] (help) from the carbon dioxide [draws arrow from C-N bond to N]. This needs to happen. Um that needs to give its electrons back to the nitrogen. Um (3.0) But I don't necessarily see any particular driving force to do that. Um (3.0) And so (3.0) with that in mind, even though I know what must happen um I question whether I know how it happens, um but given that this is just predict the product um (1.0) the-- the mechanism itself is not necessarily the most important thing um so much as that you get the-- the product out of it. So I'm just going to magically say that that happens because science.

In the end however he decides that, "the mechanism itself is not necessarily the most important thing um so much that you get the—the product out of it". Neil is more confident in his memory of the example than he is in his ability utilize chemical reasoning. This ends up leading him in the wrong direction as the removal of a BOC group would not be possible under the given reaction conditions.

This attitude was also seen in Tessa's interview when she incorrectly selects the amide nitrogen site due to similarity reasoning. While she begins to consider her choice with chemical reasoning, she becomes somewhat confused. Rather than continuing to think about the reasons why the amide would be more or less nucleophilic, she dismisses the chemical reasoning and returns to her initial similarity reasoning

Interviewer: Okay um, so how did you decided which um like nitrogen to attack with?

Tessa: I remember that – well this is the amide one [circles amide bond]. So actually you do that. Um, let's see (2.0) and I guess (1.0) hmm (2.0) I'm not sure if like a resonance argument might apply to it, which would the—that nitrogen. Well then it mm be more mm. [laugh] I just remember that you do the—the amide one.

This attitude was also reflected by Nancy during reflection questions. When asked what she could have done to better improve on Question 2, Nancy described how as much as she'd like to understand the underlying concepts, this approach has a limit and that being able to identify patterns was more important.

Interviewer: Alright what could you have done, if anything, to better prepare for this problem in your opinion?

Nancy: I think no matter how much I want to like understand the underlying concept, there's—there's a part of it that's just seeing enough of the same kind of mechanism that gets you used to the pattern, so doing more mechanism that have the Li, the Br, the ketone or and aldehyde, and then an H_3O+ workup."

While the importance of being able to categorize and recognize a variety of different problems has been seen in work looking at problem solving and is even characteristic of experts (Kraft et al., 2010; Larkin et al., 1980), Nancy seems to be expressing an attitude seen in many students that chemical reasoning alone is insufficient and that being able to recognize important parts of questions and recognize similarities to other questions is the most important. Many students also seem to overall trust their ability to make decisions based on other types of reasoning based over their ability to make decisions based on chemical reasoning.

This is surprising however when we consider the outcomes of tension points. As can be seen in Table 8.7, for question 1 tension points that are resolved with chemical reasoning lead students in the right direction the majority of the time (8 out of 9) while tension points that are resolved with other types of reasoning lead students in the wrong direction the majority of the time (7 out of 8). This pattern is supported by other analyses in this chapter that indicated that for question 1 students would have lacked the exposure to these concepts to develop the heuristics and pattern recognition skills needed to successfully use similarity, recognition, and predictive reasoning. In contrast, in the exercise-type question tension-points resolved by chemical and non-chemical reasoning both seem to lead students in the right direction the majority of the time (See Table 8.7). This is similar to what was seen at the key decision point where both predictive and chemical reasoning were effective. Chemical reasoning seems to be overall the better strategy in problem-type questions and similarly effective to other strategies in the exercise type question.

Table 8.7

Outcomes of Tension Points involving Chemical Reasoning

	Question 1 - Problem		**Question 2 - Exercise**	
Tension Point Resolution	Right Direction	Wrong Direction	Right Direction	Wrong Direction
Chemical Reasoning	8	1	4	1
Other Reasoning	1	7	4	1

Data Analysis: Strengths and Weaknesses of Chemical Reasoning

Overall the results of this study have seemed to indicate that chemical reasoning is a highly effective strategy that is necessary for making key decisions and resolving tension points in problems and which is comparable in efficacy to other strategies in exercises. Students however seem to not trust their ability to use chemical reasoning and seem to emphasize other strategies. While this suggest that instructors should encourage students to utilize chemical reasoning, there may be some limitations to this approach. This section will discuss two cases of participants who heavily used chemical reasoning and the strengths and weaknesses of their approach.

Both Rebecca and Andrea were high performing students in Organic Chemistry 1 and 2. They both received A- grades in Organic Chemistry 1 and at the end of Organic Chemistry 2 Rebecca received an A- and Andrea received an A. Both participants were also notable in that they used a lot of chemical reasoning in their problem solving. The use of chemical reasoning helped both of them to do very well on the problem-type question, and they both got the correct answer. However, both struggled more with the exercise-type question. Andrea was unable to get the correct answer and Rebecca struggled to do so and described having gotten the question incorrect on the exam.

Both Andrea and Rebecca emphasized chemical reasoning in their problem-solving approach. Their decisions were usually made using chemical reasoning and they mostly used chemical reasoning to address tension points that arose. However, each participant had a slightly different approach to their chemical reasoning. Rebecca, for example focused on identifying functional groups and their properties.

> Rebecca: One of the main things I saw in my like studying was, I guess it didn't click with until really late into my studying, that the acyl group is like super important to identify um and like what I would—what my leaving group would be…and then um identifying like what functional groups are reactive and which ones are pretty stable is helpful too.

This emphasis is reflected in Rebecca's approach to question 1 where she immediately identifies and acyl group and the leaving group and then identifies the amine as her nucleophile. A tension point appears later when she notices the presence of the amide and considers whether the amide or the amine would act as a nucleophile. Rebecca's resolution of the tension point however is again centered on chemical reasoning and she chooses the amine site because the amide nitrogen is resonance stabilized and therefore a weak nucleophile.

> Rebecca: Oh and I guess I just realized that I have an amide so I knew—okay I knew that like this nitrogen [points to amide N] wouldn't attach—attack because it is involved in um conjugation [gestures toward the amide C=O bond] and resonance stabilization so I knew that that wouldn't be a strong enough nucleophile to attack my um electrophile.

Andrea also emphasizes chemical reasoning but didn't focus on identifying reactive sites and instead considered a broader range of concepts and conditions. When asked what helped her in question 1, she mentions being familiar with protecting groups, reactive molecules, and acid-base conditions as helpful to her problem-solving.

> Interviewer: Um what did you feel like were like key components of this problem that were supported by your studying or that you like recognized because of your studying?
>
> Andrea: The BOC group definitely I recognized, that's a protecting group, boom. Also um an amine's being a good base. So I'm like this is a base we're in neutral conditions so it can have a plus or a minus charge in solution, that was definitely reinforced in my studying as well.

Similar to Rebecca, Andrea used chemical reasoning to determine next steps. She identifies the amine as the more reactive site due to the amide being conjugated. She even rationalizes the deprotonation of the amine proton as modifying the leaving group strength of the amine and then decides that sulfur would leave upon reformation of the carbonyl because it is a good leaving group.

> Andrea: Um and then I'm going to take that guy [Draws pyridine] now? Um to go ahead and protonate this guy [points to the amine in the intermediate], this amine [draws an arrow from pyridine to the H] so that way it's not gonna fall off immediately…Um so now I have a negative and a positive. This negative [points to tetrahedral intermediate] that [points to sulfur] is going to be a good leaving group so I'm gonna he—go head and make my sulfur the good leaving group.

This behavior helped both participants make good decisions and effectively resolve tension points. However even in the problem-type question, where chemical reasoning is particularly helpful, there are some limitations to a predominantly chemical reasoning approach. Both Andrea and Rebecca make a higher than average number of decisions. Andrea makes ten, Rebecca makes eight, and the average number of decisions was 6.7. Both Andrea and Rebecca also struggled to determine what the stopping point of the reaction was. They both, through undefined reasoning, feel that the reaction should continue and so consider various reactions and reactive sites before settling on the final product. Andrea considers whether the resulting sulfur anion will react with the pyridine and whether there will be subsequent reactions. Eventually she has to defer to predictive reasoning and previous practice to help her in ending the reaction.

> Andrea: So that is gonna [draws arrows from S- to the protonated pyridine]. Will that be favorable acid-base chemistry as is? Um well I'm pretty sure that it will be because I think I'm going to attack again? Or am I not? I'd probably get stuck around here? Um you know what I tried doing this at home as well, and if I attack the (p) that group I won't necessarily get that product [circles protonated pyridine in the product box]. So you know I'm just going to go ahead and put down—down this answer.

Rebecca runs into a similar dilemma, she also feels that the reaction should continue and considers whether a BOC protected amide group might be more reactive than a normal amide bond, due to the presence of the oxygen.

> Rebecca: This group is an amide—this is an amide [points to BOC protected amide] um although it has an oxygen group attached [points to tertbutyl ester group] so I don't know what that—oh I guess maybe I could attack there. Um I think in my answers I actually left it as this [points to product] um (5.0) cause I know that this is pretty stable due to resonance stabilization [points to thiocarboxylate] and I would think this is too [points to BOC protected amine]. Huh I feel like I need to attack somewhere else, that I'm not done.

Rebecca continues to consider different possible reactive sites, but in the end decides that the reaction will not continue further because she is unable to locate a good reactive site with an appropriate leaving group. Rebecca is able to use chemical reasoning to reach the correct answer but in order to do so has to consider the reactivity of various functional groups on the molecule. In contrast several other students used the protonated pyridine as a hint and assumed that the product would be negatively charged and stop at the sulfur anion for that reason.

Interestingly both Rebecca and Andrea describe having some difficulty with question 2. Both are able to recognize the relevant reactions in the question, though Rebecca tends to also rationalize her choices with chemical reasoning. Both however find themselves in a position where they have only identified one product but the hint in the product box indicates that there would be four products. Rebecca realizes at this point that the carbanion is in an allylic position and therefore there is resonance and will be multiple reactive sites. She also expresses her frustration that she did not notice this in the exam.

Rebecca: But I just realized, and I didn't do this on the exam and now I'm really upset, that I have resonance, this is an allylic position and so I can actually do resonance with my starting material so that I would make this [draws second resonance contributor]

Rebecca then reaches another tension point where she realizes that resonance only give her two products and that she is still missing two. She initially takes on a purely predictive reasoning approach and decides that the orientation of the alkene side chain could change but then realizes that the reaction created a stereocenter and that that would result in two products.

Rebecca: Um but this still confuses me because I only have two products then (7.0) and I'm also confused because—oh wait no that would make—because this is wedged [erases wedged bond] so I'm identifying that stereochem is making different product [redraws bond] Oh and—no wait, yes this is a stereoisomer [points to carbanion carbon so it can attack from front or back so that's gonna give me two products here.

Andrea however is unable to resolve the tension point between the one product that she has proposed and the four structures that the hint says will be produced. She focuses on the alkoxide carbon which is not a stereocenter. She then considers a range of different reactions and processes that would result in more products but is unable to identify any of them in the question.

Andrea: And I'm like where are my four different structures because this not a stereocenter because I have uh two methyl groups attached to it [points to two methyl groups]. So I'm like, "Did I miss something really big?"… I don't think any E1 or E2 is happening, I don't think we're switching

> around any double bonds. And, yeah I just going to write down this answer because it's the only thing I can think of and I'm running out of time on this exam.

Andrea is unable to identify a stereocenter or locate the presence of resonance in the molecule and therefore is only able to propose one structure. Interestingly however other students were able to get the correct answer without ever trying to locate stereocenters or resonance, instead these students knew that these features were implied by the four structures and simply drew them in.

Overall these two cases seem to indicate that while chemical reasoning is more thorough it is not always more efficient as students may have to consider a wide range of possible reactions and the characteristics of more functional groups and sites. In exercises students who predominantly use chemical reasoning may actually be at a disadvantage compared to students who have well developed heuristics. This also suggests that exercise-type questions may not be the best assessments because students with less chemical understanding may be able to solve them more quickly and effectively than those who understand the underlying chemical concepts more deeply

Conclusions and Implications for Instruction

Overall this chapter showed that student problem solving in problem-type questions and exercise-type questions is quite different. Students were much better able to constrain the problem space in exercise-type questions while in problem-type questions in the problem space remained quite broad. The size of the problem space meant that some participants would choose to focus on irrelevant or incorrect examples when solving their problem. This sometimes resulted in unusual solutions to the question and participants often remained focused on this pathway even when tension points arose that indicated problems with their mechanism or alternative pathways.

This work also showed that even within more standard problem-solving pathways problem-type and exercise-type questions differed in the reasoning strategies needed to solve the question. The most common reasoning strategy in question 1 was chemical reasoning while the most common reasoning strategy in question 2 was recognition reasoning. Results at key decision points showed that chemical reasoning was the most effective strategy for the problem even though it was not used very often. In contrast, for the exercise predictive reasoning, chemical reasoning, and combinations of chemical and predictive reasoning were all very effective strategies. This indicated that problem-type questions seemed to require the use of chemical reasoning while exercise-type questions could be solved with a variety of strategies.

This result was emphasized when looking at the resolution of tension points in each question. In the problem-type question chemical reasoning resulted in students making a choice that helped them move in the right direction the majority of the time while other strategies moved students in the wrong direction the majority of the time. In the exercise-type question, chemical reasoning and other types of reasoning were both effective ways to resolve tension points. Overall these results also showed that while chemical reasoning is a consistently effective strategy it is not widely taken up by students. Students in interviews seemed to undervalue chemical reasoning as a strategy and consider other strategies more valuable and more convinced. Some of this might be because, as Rebecca and Andrea's cases showed, chemical reasoning can be a less efficient strategy and because in exercise type problems it may be less effective than a well-developed heuristic.

The results have several implications for the instruction of organic chemistry. The first is that this work shows that students utilize various reasoning strategies over the course of problem-solving and that these strategies will sometimes be in conflict, particularly in problem-type questions. It also shows that some of reasoning strategies are more effective than others. It may be helpful for instructors to model these reasoning strategies and to emphasize chemical reasoning. It may also be helpful for instructors to discuss tension points and emphasize the use of chemical reasoning to resolve them, since it is effective in both exercises and problems. This is important because students often dismiss chemical reasoning in tension points. In problems they will often preferentially focus in on fragmented and incorrect memories of previous examples. Emphasizing tension points in instruction may also be helpful because in situations where students can't resolve tension points, they often give up.

These results also have implications for the development of exam items. They indicate that problem-type and exercise-type questions require the use of different reasoning strategies and that only problems seem to consistently require the use of chemical reasoning. If instructors want to ensure that their items are measuring students understanding of chemical concepts, they should be cautious about the use of too many exercise questions since they can be solved without the use of chemical reasoning. High levels of exercise-type questions may not reward chemical reasoning thus further discourage the use of this reasoning strategy.

There are some limitations to this study, the most notable of which is the small sample size. Future work would ideally expand this analysis to a larger number of participants, particularly those with lower grades in the course. Future work would also include more comparisons of problems and exercises to try to generalize results about reasoning strategies and problem space size to a broader set of questions. The other limitation is that students had previously seen the question on the exam. While this allowed for studying of exam questions within a course-context, students were more familiar with these questions than they would have been with questions designed specifically for an interview. Student interviews indicated that most students did not remember how they solved the questions, and in some cases did not remember the questions themselves. This was supported by analysis of student answers on the exam that indicated that students were not consistent with their answers between the exam and the interview. Another limitation is the nature of think-aloud interviews which does not necessarily capture all of student thinking, particularly for students who may not instinctively verbalize their thinking when problem-solving. This may mean that some of the participants' decisions were not described out-loud or that the reasoning behind them was not completely vocalized. This was somewhat mitigated by the use of reflective questions though it was not possible to analyze every student decision in the moment.

However even with these limitations that results still indicate that a fine-grained analysis of student problem solving can shed light on important characteristics of student problem-solving in organic chemistry particularly the differences that occur between problem-type and exercise-type exercises. They also suggest that instructors should be aware of these characteristics (particularly problem-solving phases and reasoning strategies) when they instruct students and when they develop their assessments.

Chapter 9: Conclusions

The purpose of this dissertation was to explore how study-behaviors and problem-solving strategies influence student performance in Organic Chemistry 2, a second semester organic chemistry course for non-majors. This work in particular, however, was interested in studying how these relationships might differ across different types of questions that appeared in exams. This dissertation, therefore, considered differences between common categories of organic chemistry questions (mechanism and predict-the-product) and differences in the relationship the questions have to the curriculum (problem-type and exercise-type).

This work began by showing that four main study behaviors were associated with overall exam performance in Organic Chemistry 2 (studying more than 10 hours, writing one's own thoughts/ideas when taking notes, predominantly testing one's understanding when doing practice problems, and passive engagement in group work). However subsequent analyses indicated that these results were not consistent across different types of questions. For example, Chapter 6 showed that different categories of questions (mechanism and PTP) were associated with very different study behaviors. Chapter 7, showed that even within a type of question differences can be significant. Within the category of predict-the-product questions problem-type and exercise-type questions were associated with very different study behaviors.

Having identified notable differences in the forms of preparation that lead to high outcomes on problems and exercises I was then interested in how student problem-solving might influence student performance on problem and exercise type questions. Rather than focus on overarching reasoning strategies, this chapter explored the various reasoning strategies that students utilize as they solve a given question, acknowledging that students will likely use multiple strategies which will sometimes come into conflict. Three main phases of problem solving were identified: Brainstorming, Decision-Points and Tension points. Students were also observed to use a range of reasoning strategies during these phases, though five main categories were identified: Recognition, Similarity, Undefined, Predictive, and Chemical reasoning. Results showed that student problem-solving on problem-type and exercise-type questions were fairly different and that students are better able to constrain the problem space in problems versus exercises. Other results indicated that chemical reasoning was in general an effective strategy in organic chemistry problem solving, but that in exercises recognition and predictive reasoning was equally effective. Results also showed that students tend to underuse and undervalue chemical reasoning and will often pick other less effective strategies over it, particularly at tension points.

These results have significant implications for organic chemistry instruction. Results from Chapters 4-7 indicate that exams are not homogenous and that there is not a single set of study behaviors that can help students across all types of questions. Instead, when advising students, instructors should consider how students are doing on specific types of problems. For example, if a student is struggling with mechanism question increasing their collaboration with peers while in lecture would be helpful. However, if a student is struggling with predict-the-product questions encouraging them to consistently work on practice problems and think critically about their understanding and how ideas connect would be more effective. Instructors should also consider these differences in their exam designs. Ensuring that exams have a wide range of questions of different types and with different relationships to the curriculum would ensure that they are not strongly preferring one type of study behavior over another. Instructors should also consider regularly assessing students study behaviors and its impact on exam scores and sub-scores to ensure that determine the effectiveness of the resources they are providing students and to ensure that their exams are not associated with more passive types of engagement.

Chapter 8 also has considerable implications for organic chemistry instructors. It is important that instructors be aware that students are utilizing multiple types of reasoning when they problem solve and that these reasoning strategies can come into conflict. It is also important that instructors be aware that students will often undervalue chemical reasoning even though it is the most consistently effective reasoning strategy across problems and exercises.

Instructors might consider talking with students explicitly about the decisions that are made when problem solving and the different forms of reasoning that they might use to make decisions. By emphasizing different ways of solving questions and even modeling potential mistakes instructors could help students be more aware of their problem-solving and the decision that they make. These discussions should encourage students to consider chemical reasoning, but instructors should be aware that chemical reasoning is not always the most efficient strategy and that students may want to pair it with other reasoning strategies as well. Instructors should also discuss tension points and effective ways to resolve them, emphasizing the general strength of chemical reasoning. This is particularly important because students who were unable to reach solutions usually were unable to resolve a tension point.

Instructors should also be aware of these differences when designing assessments. Exercise-type questions can be solved without the use of chemical reasoning and so instructors should be wary of using them too often. Instructors who wish to test students deeper chemical understanding should also be aware the questions that are very similar to what students have seen in lectures and on problem sets will often not measure students understanding of chemical concepts since students are able to solve those problem with predictive reasoning strategies alone.

Organic Chemistry has a large influence on student persistence in STEM and medical fields. In order to better support students it is important to understand what strategies, both in studying and problem-solving help students to perform well. As I have shown in this dissertation, however, this process involves considering the nature of the exams being used to assess students. In considering the categories of questions being used and their relationship to the curriculum we are better able to identify the study behaviors and problem-solving strategies that can help students learn deeply and excel in organic chemistry.

References

Austin, A. C., Ben-Daat, H., Zhu, M., Atkinson, R., Barrows, N., & Gould, I. R. (2015). Measuring student performance in general organic chemistry. *Chemistry Education Research and Practice*, *16*(1), 168–178. https://doi.org/10.1039/c4rp00208c

Barr, D. A., Gonzalez, M. E., & Wanat, S. F. (2008). The leaky pipeline: Factors associated with early decline in interest in premedical studies among underrepresented minority undergraduate students. *Academic Medicine*, *83*(5), 503–511. https://doi.org/10.1097/ACM.0b013e31816bda16

Barr, D. A., Matsui, J., Wanat, S. F., & Gonzalez, M. E. (2010). Chemistry courses as the turning point for premedical students. *Advances in Health Sciences Education*, *15*(1), 45–54. https://doi.org/10.1007/s10459-009-9165-3

Bhattacharyya, G., & Bodner, G. M. (2005). "It Gets Me to the Product": How Students Propose Organic Mechanisms. *Journal of Chemical Education*, *82*(9), 1402. https://doi.org/10.1021/ed082p1402

Black, A. E., & Deci, E. L. (2000). The Effects of Instructors' Autonomy Support and Students' Autonomous Motivation on Learning Organic Chemistry: A Self-Determination Theory Perspective. *Science Education*, *84*(6), 740–756. https://doi.org/10.1002/1098-237X(200011)84:6<740::AID-SCE4>3.0.CO;2-3

Bodé, N. E., & Flynn, A. B. (2016). Strategies of Successful Synthesis Solutions: Mapping, Mechanisms, and More. *Journal of Chemical Education*, *93*(4), 593–604. https://doi.org/10.1021/acs.jchemed.5b00900

Bodner, G. M. (1987). The role of algorithms in teaching problem solving. *Journal of Chemical Education*, *64*(6), 513. https://doi.org/10.1021/ed064p513

Bodner, G. M., & Domin, D. S. (2000). Mental models: The role of representations in problem solving in chemistry. *Chemistry Education: Research and Practice*, *4*(1), 24–30. https://doi.org/10.1038/nrc1891

Bodner, G. M., & Herron, J. D. (2002). Problem-Solving in Chemistry. In J. K. Gilbert, O. De Jong, R. Justi, D. F. Treagust, & J. H. Van Driel (Eds.), *Chemical Education: Towards Research-based Practice* (p. 235). Kluwer Academic Publishers.

Bowen, C. W., & Bodner, G. M. (1991). Problem-solving processes used by students in organic synthesis. *International Journal of Science Education*, *13*(2), 143–158. https://doi.org/10.1080/0950069910130202

Bransford, J. D., Brown, A., & Cocking, R. R. (2000). *How People Learn*. *How People Learn*. Washington, D.C.: National Academies Press. https://doi.org/10.17226/9853

Cartrette, D. P., & Bodner, G. M. (2010). Non-mathematical problem solving in organic chemistry. *Journal of Research in Science Teaching*, *47*(6), 643–660. https://doi.org/10.1002/tea.20306

Christian, K., & Talanquer, V. (2012). Modes of reasoning in self-initiated study groups in chemistry. *Chemistry Education Research and Practice*, *13*(3), 286–295. https://doi.org/10.1039/c2rp20010d

Cracolice, M. S., Deming, J. C., & Ehlert, B. (2008). Concept Learning versus Problem Solving: A

Cognitive Difference. *Journal of Chemical Education*, *85*(6), 873. https://doi.org/10.1021/ed085p873
Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and Conducting Mixed Methods Research* (2nd Editio). Thousand Oaks, CA: SAGE Publications.
DeCocq, V., & Bhattacharyya, G. (2018). 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. https://doi.org/10.1039/c8rp00214b
Entwistle, N., & McCune, V. (2004). The conceptual bases of study strategy inventories. *Educational Psychology Review*, *16*(4), 325–345. https://doi.org/10.1007/s10648-004-0003-0
Flynn, A. B. (2014). How do students work through organic synthesis learning activities? *Chemistry Education Research and Practice*, *15*(4), 747–762. https://doi.org/10.1039/c4rp00143e
Hayes, J. (1980). *The Complete Problem Solver*. Philadelphia: Franklin Institute.
Holme, T., Bretz, S. L., Cooper, M., Lewis, J., Paek, P., Pienta, N., … Towns, M. (2010). Enhancing the role of assessment in curriculum reform in chemistry. *Chemistry Education Research and Practice*, *11*(2), 92–97. https://doi.org/10.1039/c005352j
Holme, T., & Murphy, K. (2012). The ACS Exams Institute Undergraduate Chemistry Anchoring Concepts Content Map I: General Chemistry. *Journal of Chemical Education*, *89*(6), 721–723. https://doi.org/10.1021/ed300050q
Hurtado, S., Han, J. C., Sáenz, V. B., Espinosa, L. L., Cabrera, N. L., & Cerna, O. S. (2007). Predicting transition and adjustment to college: Biomedical and behavioral science aspirants' and minority students' first year of college. *Research in Higher Education*, *48*(7), 841–887. https://doi.org/10.1007/s11162-007-9051-x
Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradign Whose Time Has Come. *Educational Researcher*, *33*(7), 14–26. https://doi.org/10.2307/3700093
Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, *12*(1), 1–48. https://doi.org/10.1016/0364-0213(88)90007-9
Kraft, A., Strickland, A. M., & Bhattacharyya, G. (2010). Reasonable reasoning: Multi-variate problem-solving in organic chemistry. *Chemistry Education Research and Practice*, *11*(4), 281–292. https://doi.org/10.1039/c0rp90003f
Larkin, J., Mcdermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and Novice Performance in Solving Physics Problems. *Science*, *208*(4450), 1335–1342. Retrieved from http://www.jstor.org/stable/1684057
Linn, M. C., Lee, H., Tinker, R., Husic, F., & Chiu, J. L. (2006). Teaching and Assessing Knowledge Integration in Science Teaching and Assessing Knowledge Integration Supporting Online Material. *Science*, *313*, 1049–1051. https://doi.org/10.1126/science.1131408
Lovecchio, K., & Dundes, L. (2002). Premed Survival: Understanding the Culling Process in Premed... : Academic Medicine. *Academic Medicine*, *77*(7), 719–724. Retrieved from http://journals.lww.com/academicmedicine/Fulltext/2002/07000/Premed_Survival__Understanding_the_Culling_Process.16.aspx
Lynch, D. J., & Trujillo, H. (2011). Motivational beliefs and learning strategies in organic

chemistry. *International Journal of Science and Mathematics Education*, *9*(6), 1351–1365. https://doi.org/10.1007/s10763-010-9264-x

MARTON, F., & SÄLJÖ, R. (1976). ON QUALITATIVE DIFFERENCES IN LEARNING: I-OUTCOME AND PROCESS*. *British Journal of Educational Psychology*, *46*(1), 4–11. https://doi.org/10.1111/j.2044-8279.1976.tb02980.x

Miller, D. I., & Wai, J. (2015). The bachelor's to Ph.D. STEM pipeline no longer leaks more women than men: A 30-year analysis. *Frontiers in Psychology*, *6*(FEB), 1–10. https://doi.org/10.3389/fpsyg.2015.00037

Murphy, K., Holme, T., Zenisky, A., Caruthers, H., & Knaus, K. (2012). Building the ACS exams anchoring concept content map for undergraduate chemistry. *Journal of Chemical Education*, *89*(6), 715–720. https://doi.org/10.1021/ed300049w

Nakhleh, M. B., & Mitchell, R. C. (1993). Concept learning versus problem solving: There is a difference. *Journal of Chemical Education*, *70*(3), 190. https://doi.org/10.1021/ed070p190

Nurrenbern, S. C., & Pickering, M. (1987). Concept learning versus problem solving: Is there a difference? *Journal of Chemical Education*, *64*(6), 508. https://doi.org/10.1021/jacs.6b02172

Paulson, D. R. (1999). Active Learning and Cooperative Learning in the Organic Chemistry Lecture Class. *Journal of Chemical Education*, *76*(8), 1136. https://doi.org/10.1021/ed076p1136

Pintrich, P. R. (2004). A Conceptual Framework for Assessing Motivation and Self-Regulated Learning in College Students Author (s): Paul R . Pintrich Source : Educational Psychology Review , Vol . 16 , No . 4 , Measuring Studying and Learning in Higher Education — Conceptual a. *Springer*, *16*(4), 385–407. https://doi.org/10.1007/s10648-004-0006-x

Polya, G. (1945). *How to Solve It: A New ASpect of Mathematical Method*. Princeton, NJ: Princeton University.

Pribyl, J. R., & Bodner, G. M. (1987). Spatial Ability and its Role in Organic Chemistry: A study of four organic chemistry courses. *Journal of Research in Science Teaching*, *24*(3), 229–240. Retrieved from papers3://publication/uuid/3C0BCB9D-9230-45D0-867D-3083EACF00EE

R.Anderson, John. (2009). Problem solving and learning. *AIP Conference Proceedings*, *1140*(1972), 183–197. https://doi.org/10.1063/1.3183522

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and Longitudinal Modeling Using Stata Volume 1: Continuous Responses* (Third Edit). College Stattion, TX: Stata Press.

Raker, J. R., Holme, T. A., & Murphy, K. L. (2013). The ACS Exams Institute Undergraduate Chemistry Anchoring Concepts Content Map II: Organic Chemistry. *Journal of Chemical Education*, *90*, 1443–1445. https://doi.org/10.1021/acs.jchemed.7b00531

Schoenfeld, A. H. (1988). When Good Teaching Leads to Bad Results: The Disasters of "Well-Taught" Mathematics Courses. *Educational Psychologist*, *23*(2), 145–166. https://doi.org/10.1207/s15326985ep2302_5

Seymour, E., & Hewitt, N. M. (1997). *Talking About Leaving: Why Undergraduates Leave The Sciences*. Boulder, CO: Westview.

Sinapuelas, M. L. S., & Stacy, A. M. (2015). The relationship between student success in introductory university chemistry and approaches to learning outside of the classroom. *Journal of Research in Science Teaching*, *52*(6), 790–815. https://doi.org/10.1002/tea.21215

Solomon, P. (1995). *The think aloud method: A practical guide to modelling cognitive processes*. *Information Processing & Management* (Vol. 31). https://doi.org/10.1016/0306-4573(95)90031-4

Stern, L., & Ahlgren, A. (2002). Analysis of students' assessments in middle school curriculum materials: Aiming precisely at benchmarks and standards. *Journal of Research in Science Teaching*, *39*(9), 889–910. https://doi.org/10.1002/tea.10050

Szu, E., N, Agopal, K., Shavelson, R. J., Lopez, E. J., Penn, J. H., … Hill, G. W. (2011). Understanding Academic Performance in Organic Chemistry. *Journal of Chemical Education*, *88*(9), 1238–1242. https://doi.org/10.1021/ed900067m

Tien, L. T., Roth, V., & Kampmeier, J. A. (2002). Implementation of a peer-led team learning instructional approach in an undergraduate organic chemistry course. *Journal of Research in Science Teaching*, *39*(7), 606–632. https://doi.org/10.1002/tea.10038

Webber, D. M., & 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. research-article. https://doi.org/10.1021/acs.jchemed.8b00158

Zoller, U., Dori, Y. J., & Lubezky, A. (2002). Algorithmic, LOCS and HOCS (chemistry) exam questions: Performance and attitudes of college students. *International Journal of Science Education*, *24*(2), 185–203. https://doi.org/10.1080/09500690110049060

Zoller, U., Lubezky, A., Nakhleh, M. B., Tessier, B., & Dori, Y. J. (1995). Success on Algorithmic and LOCS vs. Conceptual Chemistry Exam Questions. *Journal of Chemical Education*, *72*(11), 987. https://doi.org/10.1021/ed072p987

Zusho, A., Pintrich, P. R., & Coppola, B. (2003). Skill and will: The role of motivation and cognition in the learning of college chemistry. *International Journal of Science Education*, *25*(9), 1081–1094. https://doi.org/10.1080/0950069032000052207

Appendix A: Survey Protocol (Time 3)

Section 1: Course Background
Q1 What were your motivations for taking [Organic Chemistry 2]?

Q2 Had you previously taken [Organic Chemistry 2] at [University of the West Coast]?

- Yes
- No

If answer to Q2 was yes:
Q3 Please indicate in which semester(s) you took [Organic Chemistry 2] (check all that apply):

- Fall 2014
- Spring 2015
- Summer 2015
- Fall 2015
- Spring 2016
- Summer 2016
- Fall 2016
- Spring 2017
- Summer 2017
- Fall 2017
- Other (please specify): _________________________

Q4 Have you previously taken [Organic Chemistry 1] at [University of the West Coast]?

- Yes
- No

If answer to Q3 was yes:
Q5 Please indicate in which semester(s) you took [Organic Chemistry 1] (check all that apply):

- Fall 2014
- Spring 2015
- Summer 2015
- Fall 2015
- Spring 2016
- Summer 2016
- Fall 2016
- Spring 2017
- Summer 2017
- Fall 2017
- Other (please specify): _________________________

If answer to Q3 was no:
Q6 Did you complete another organic chemistry course instead of [Organic Chemistry 1]?

- Yes
- No

If answer to Q3B was Yes:
Q7 Please indicate the name of the course and at what institution is was offered.

If answer to Q3B was Yes:
Q8 Please indicate in which semester(s) you completed this course (check all that apply):

- Fall 2014
- Spring 2015
- Summer 2015
- Fall 2015
- Spring 2016
- Summer 2016
- Fall 2016
- Spring 2017
- Summer 2017
- Fall 2017
- Other (please specify): ___________________________

Section 2: In-Class Study Behaviors
Q9 Which lecture time did you normally attend?

- Morning (Tu/Thurs 8-9:30 am)
- Afternoon (Tu/Thurs 3:30-5 pm)

Q10 the last week of class how often did you attend lecture?

- I attended both lectures
- I attended one of the lectures
- I did not attend lecture

Q11 During the last week of class what did you do during lecture (check all that apply)?

- Taking notes during lecture
- Completing in-class problems/worksheets
- Answering iClicker questions
- Volunteering to answer questions posed by the instructor
- Asking questions of the instructor
- Collaborating/Talking with neighbors

Q12 During the last week of class what kind of notes did you take in [Organic Chemistry 2] (check all that apply)?

- Wrote down what the instructor wrote down verbatim
- Wrote down what the instructor said verbatim
- Translated what the instructor said into my own words
- Wrote down my own thoughts/insights

Q13 During the last week of class what topics did you include in your notes (check all that apply)?

- Recorded key words and ideas I'd need to remember
- Wrote down examples of reactions
- Wrote down underlying concepts
- Recorded problem solving strategies

Q14 Reflecting on the previous two questions, why do you take this type of notes?

Q15 How do you take notes? (check all that apply)

- Paper and writing utensil
- Laptop computer (non-tablet)
- Tablet computer

Q16 How do you use in-class problems/worksheets (check all that apply)?

- Wait until the instructor goes over the problem(s) and write down the answer(s)
- Work on the problem, then revise the solution when the instructor goes over the problem.
- Work alone to solve the problem(s).
- Work with a neighbor/small group to solve the problem(s).
- Ask the instructor/GSI questions about solving the problem(s).

Q17 The last time I worked with others on an in-class activity/worksheet:

- I mostly asked questions.
- I mostly answered questions
- I both asked and answered questions
- I mostly listened

Section 3: Study Hours and Additional Instructional Resources

Q18 During [Reading] week how many hours did you study (including readings and problem sets but not including work from [Organic Chemistry 1] lab)?

- < 1 hour
- 1-4 hours
- 4-7 hours
- 7-10 hours
- >10 hours

Q19 Please indicate which of the following forms of additional instruction you used during [Reading] week (check all that apply):

- Review sessions held by the head GSI
- GSI office hours
- Instructor office hours
- SLC drop-in tutoring
- Reading/Writing on Piazza message board
- Emailing the instructor/GSI about the material
- Private tutoring

Q20 During [Reading] week how did you use the Piazza message boards? (check all that apply)

- I didn't
- I read responses to the questions of others
- I asked questions
- I had my questions answered
- I answered questions posted by others

Section 4: Group-work Environments

Q21 Please indicate for the following how frequently you participated in group work during [Reading] week:

	Never	10 min - 1 hr	1 hr - 4 hr	4+ hr
Studied with a buddy	○	○	○	○
Studied with a group of friends/peers	○	○	○	○
SLC study group	○	○	○	○
BSP study group	○	○	○	○
Studied with a group in my residence or social organization	○	○	○	○

If in Q21 response was > Never
Q22 How did the group help you learn?

If in Q21 response was > Never
Q23 In our last group meeting:

- I mostly asked questions
- I mostly answered questions
- I both asked and answered questions
- I mostly listened

Section 5: Using Online or Written Resources

The following section asks about resources you could have utilized outside of class in which you **reviewed written or online resources.**

Q24 During [Reading] week, I:

	Never	10 min - 1 hr	1 hr - 4 hr	4+ hr
Read the textbook	◯	◯	◯	◯
Reviewed lecture notes	◯	◯	◯	◯
Reviewed midterms/quizzes	◯	◯	◯	◯
Reviewed in-class problems/worksheets	◯	◯	◯	◯
Read other textbooks or wikis	◯	◯	◯	◯
Watched related videos	◯	◯	◯	◯

Q25 When reviewing written or online resources I **predominantly** try to (select up to three responses):

- Memorize key words, concepts, molecules, and reactions
- Review problem types that might appear on a quiz/exam
- Test my understanding of the material
- Build connections among ideas
- Build connections between the material and the real world
- Identify underlying concepts
- Review problem solving approaches

If in Q24 Read the Textbook > Never
Q26 Would you read the textbook before or after the relevant lecture?

- Before the relevant lecture
- After the relevant lecture
- Both before and after the relevant lecture

Q27 Why do you rewrite your lecture notes? (check all that apply)

- I don't
- To make them legible
- To translate them into my own words
- To help me review the material
- To fill in gaps or identify gaps in my understanding

Q28 Which of the following study materials did you develop to help with learning the material (check all that apply):

- Flash cards
- Reaction lists
- Table/charts of the material
- Flow charts
- Outlines/Review sheets of key concepts

Section 6: Practice Problems

Q29 My activities during Reading week included:

	Never	10 min - 1 hr	1 hr - 4 hr	4+ hr
Problem sets	○	○	○	○
Practice exams	○	○	○	○
Textbook Problems	○	○	○	○
Online problems (unaffiliated with the class)	○	○	○	○
SLC mock midterm	○	○	○	○

Q30 How do you use answer keys for practice problems?

- I review the answer key along with the problem
- I check the answer key after each problem
- I check the answer key after each page
- I check the answer key after I try all or most of the problems
- I ignore the answer key

Q31 When working on practice problems I **predominantly** try to (select up to three responses):

- Memorize key words, concepts, molecules, and reactions
- Review problem types that might appear on quiz/exam
- Test my understanding of the material
- Build connections among ideas
- Build connections between the material and the real world
- Identify underlying concepts
- Review problem solving approaches

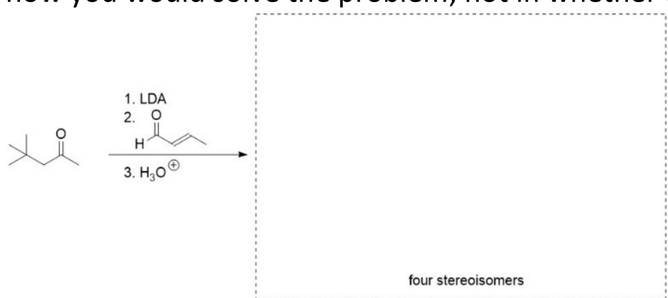
Section 7: Reflection Questions

Q32 Describe how you studied/are studying for the final exam. Why have you found this form of studying helpful to you?

Q33 If you've changed the way you study since the second midterm what didn't work for you in the past and why?

Q34 Describe how you would approach solving the following problem. We are interested in how you would solve the problem; not in whether or not your answer would be correct.

Q35 Describe how you would approach solving the following problem. We are interested in how you would solve the problem; not in whether or not your answer would be correct.



Section 8: Demographics & Background

Q36 Did you transfer to [University of the West Coast] from another college or university?

- Yes
- No

Q37 What is your major or intended major?

- Bioengineering
- Chemical Biology
- Chemical Engineering
- Chemistry
- Civil Engineering
- Computer Science
- Environmental/Natural Science
- Humanities
- Life Science/Biology
- Molecular and Cell Biology (MCB)
- Molecular Evolutionary Biology
- Mathematics
- Mechanical Engineering
- Nutrition Science
- Other Engineering
- Physical Science
- Public Health
- Social Science
- Other: ____

Q38 Do you consider yourself pre-med?

- Yes
- No

Q39 How many college credit units did you take this semester?

Q40 How many hours per week, on average, did you work for pay? (please round to the nearest whole number)

Q41 Please indicate your final grade in your previous organic chemistry course ([Organic Chemistry 1] or equivalent). If you took the course multiple times please indicate the grade you received the last time you completed the course.

- A+
- A
- A-
- B+
- B
- B-
- C+
- C
- C-
- D
- F
- P
- NP
- Other (please specify): ________________________
- Decline to state

Q42 What is the highest level of formal education obtained either of your parents/guardians?

- Did not complete high school
- High school graduate
- Postsecondary school other than college
- Some college
- College degree
- Some graduate school
- Graduate degree
- Not sure
- Decline to state

Q43 Are you: (Mark **all** that apply)

- White/Caucasian
- African American/Black
- American Indian/Alaska Native
- Middle Eastern/North African (e.g., Moroccan, Egyptian, Saudi Arabian, Iranian)
- East Asian (e.g., Chinese, Japanese, Korean, Taiwanese)
- Filipino
- Southeast Asian (e.g., Cambodian, Vietnamese, Hmong)
- South Asian (e.g., Indian, Pakistani, Nepalese, Sri Lankan)
- Other Asian
- Native Hawaiian/Pacific Islander
- Mexican American/Chicano
- Puerto Rican
- Other Latino
- Other (please specify): ____
- Decline to state

Q44 What is your gender?

- Male
- Female
- Female to male transgender
- Male to female transgender
- Not sure
- Decline to state
- Other: ____

Q45 We appreciate your feedback and ask for your student ID to verify your enrollment in the course and provide you with extra credit points. Before this data is reported, your student ID will be removed. Thank you for your participation. ________________________________

Appendix B: Interview Protocol (Time 1)

Introductory questions:

1. Overall how would you describe your experience in [Organic Chemistry 2] so far?
    a. Could you talk me through some of the high and low points?
2. How did you feel about how you did on the first [Organic Chemistry 2] midterm?
3. Could you describe your approach to studying for the [Organic Chemistry 2] midterm?
    a. Could you describe what types of activities you normally did during lecture?
    b. If they bring up note-taking: What type of notes did you take?
    c. If they bring up group work: What was your role in the group? When you work in a group what are you looking for?
    d. If they bring up office hours: What do you get out of office hours? Are you able to ask questions, what kinds?
4. What kinds of questions do you ask yourself? Do you ask others?
5. How did you develop these approaches?
    a. Were there any strategies you felt particularly encouraged or discouraged to use?
    b. Who did you hear about study strategies from?

Follow up prompts:
What do you mean by that?
Would you describe it to me?
How do decide what __ to look for?
How do you determine __ (what you don't know, what to take notes on etc).
What type of __? (ie. Patterns)

Think-aloud:

Introduction script:
Thank you so much, now I am interested in how students approach the questions on exams in organic chemistry. What I'm going to have you do is work through a few predict-the-products organic chemistry problems from your midterm and from other course materials, and I want you to vocalize your thoughts as you have them to the best of your ability. I'll prod you if your stop talking for too long. Also, I should point out that you're not explaining it to me or anybody else. We're trying to get at the best approximation of the thoughts you'd have working on the problem alone. After you're done I'll ask you a few follow-up questions.

1. Do you have any questions for me?

Here is question 1:
Please describe out loud what you are thinking as you solve it now. Please try to keep talking as you approach the problem. All your ideas are of interest.

1. Some other students also got to this alternative product (draw product). How do you think they might have gotten to this product? *Show correct product if answer is incorrect and incorrect product if answer is incorrect*
2. What did you do that helped you prepare for this problem?
3. What could you have done, if anything, to prepare better for this problem?

Prompts (if participants get stuck):
I am just interested in how you solve the problem; not in the right answer to the problem.
What are you thinking now?
Keep describing your thinking.
Keep talking out loud
If they answer too quickly: Why are those the answer?

Appendix C: Observation Protocols

Outline for Table Format

Topic	Sub-category	Count	Description
Pedagogical Choices	Group Work iClicker Questions to the crowd Solve problems out loud/on board Other		 1. Quick Check-ins (Yes? Good?) 1.
Encouragements for in-class behavior	Asking questions Solve problems in class Talk with neighbor Other		
Encouragement for out-of-class behavior	Textbook Talking to previous students Reviewing problem sets/lecture notes Doing problems from lecture notes at home Doing problem sets Other		
Topics covered in class			
Trends/Comparisons Emphasized			
Types of questions being asked by students			
Callbacks	To [Organic Chemistry 1] Materials The previous [Organic Chemistry 2] Material To General Chemistry Other		
Connections to Biology			
Reasoning Strategies	Case-Based Independent Rules Rules that summarize theory Model-based Definition/Explicit Memorization Other		
Exam/Coursework Strategies	Things to do Things to ignore		

Other Topics:

Detailed descriptions of predict-the-product questions:

Outline of Annotated Chronology (with a partial example)

Concept	How it's explained/Student Questions	Type of reasoning (Examples: Memorization, rules based, case based, model based, exam/coursework strategies, other)	Type of Active Learning/ Out of Class Behavior Encouraged
	- Make sure to note callbacks to previous material (from the course or other courses) - Make sure to note explicit connections to biology - Emphasize predict the product questions		
Administrative	Instructor hands out worksheets before the class start time.		
Lithium Aluminum	Call back to last lecture		Call back
Hydride reacts with everything	Aluminum neutralizes charge on oxygen or nitrogen and allows it to potentially work as a leaving group		
LAH reaction with amide – selectivity of this reaction	Discusses reaction with amide.		Work on problem
	Has students work with the mechanism Use aluminum as a hydride donor and to neutralize any negative charges LAH is also a strong base so the first step with be to produce H2 gas Aiming for a key intermediate where the selectivity happens		Talk to GSIs
	Also we have GSIs here every day who would love to talk to you guys. They were telling me they only get iclicker questions and would love chemistry ones		
	#Students nearby conversing (not about chemistry)	Trends	

Aluminum hydrides are very strong bases. This is similar to what we say with a Grignard. Hydride is going to act as a base whether we like it or not.

Rule

Appendix D: Tables and Graphs

Table D.1: Frequency Table of Categorical Variables

	Occasion 1 (N = 357)	Occasion 2 (N = 345)	Occasion 3 (N = 382)
Study Hours			
<4 hours	36 (10%)	34 (10%)	48 (13%)
4-7 hours	41 (11%)	40 (12%)	46 (12%)
7-10 hours	82 (23%)	78 (23%)	66 (17%)
10 + hours	198 (54%)	193 (56%)	222 (58%)
In-Class Study Behaviors			
Collaborated/Talked with neighbors during lecture			
Did not collaborate/talk with neighbors	84 (24%)	108 (31%)	153 (40%)
Collaborated/Talked with neighbors	273 (76%)	237 (69%)	229 (60%)
Wrote down own thoughts/words when taking notes			
Did not write down own thoughts/words	130 (36%)	117 (34%)	155 (41%)
Wrote down own thoughts/words	227 (64%)	228 (66%)	227 (59%)
Worked on problem, then revised the solution when the instructor went over the problem			
Did not work on problem, then revise the solution	95 (27%)	79 (23%)	120 (31%)
Worked on problem, then revised the solution	262 (73%)	266 (77%)	262 (69%)
Use of additional Instructional Resources			
Review Session			
Did not attend review session	276 (77%)	274 (79%)	263 (69%)
Attended review session	81 (23%)	71 (21%)	119 (31%)
Office Hours			
Did not attend office hours	286 (80%)	263 (76%)	322 (84%)
Attended office hours	71 (20%)	82 (24%)	60 (16%)
Piazza			
Did not use Piazza	89 (25%)	81 (23%)	105 (27%)
Used Piazza passively	189 (53%)	192 (56%)	218 (57%)
Used Piazza actively	79 (22%)	72 (21%)	59 (15%)
Use of written/online resources			
Memorized key words, concepts, molecules, and reactions			
Did not memorize key words, concepts &reactions	135 (38%)	153 (44%)	161 (42%)
Memorized key words, concepts & reactions	225 (62%)	192 (56%)	221 (58%)
Reviewed problem types that might appear on quiz/exam			
Did not review problem types	100 (28%)	96 (28%)	120 (31%)
Reviewed problem types	257 (72%)	249 (72%)	262 (69%)
Tested understanding of material			
Did not test understanding of material	186 (52%)	165 (48%)	171 (45%)
Tested understanding of material	171 (48%)	180 (52%)	211 (55%)
Built connections among ideas			
Did not build connections among ideas	272 (76%)	247 (72%)	301 (79%)
Built connections among ideas	85 (24%)	98 (28%)	81 (21%)
Identified underlying concepts			
Did not identify underlying concepts	235 (66%)	222 (64%)	247 (65%)
Identified underlying concepts	122 (34%)	123 (36%)	135 (35%)
Reviewed problem-solving approaches			
Did not review problem-solving approaches	219 (61%)	228 (66%)	240 (63%)
Reviewed problem-solving approaches	138 (39%)	117 (34%)	142 (37%)
Use of practice problems			
Memorized key words, concepts, molecules, and reactions			
Did not memorize key words, concepts &reactions	168 (47%)	181 (52%)	174 (46%)
Memorized key words, concepts & reactions	189 (53%)	164 (48%)	208 (54%)
Reviewed problem types that might appear on quiz/exam			

Did not review problem types	113 (32%)	103 (30%)	137 (36%)
Reviewed problem types	244 (68%)	242 (70%)	245 (64%)
Tested understanding of material			
Did not test understanding of material	149 (42%)	143 (41%)	161 (42%)
Tested understanding of material	208 (58%)	202 (59%)	221 (58%)
Built connections among ideas			
Did not build connections among ideas	307 (86%)	300 (87%)	348 (91%)
Built connections among ideas	50 (14%)	45 (13%)	34 (9%)
Identified underlying concepts			
Did not identify underlying concepts	238 (67%)	219 (63%)	258 (68%)
Identified underlying concepts	119 (33%)	126 (37%)	124 (32%)
Reviewed problem-solving approaches			
Did not review problem-solving approaches	217 (61%)	217 (63%)	223 (58%)
Reviewed problem-solving approaches	140 (39%)	128 (37%)	159 (42%)
Group Study			
Did not engage in group work	99 (28%)	109 (32%)	146 (38%)
Passively engaged in groupwork	28 (8%)	32 (9%)	30 (8%)
Actively engaged in groupwork	230 (64%)	204 (59%)	206 (54%)

Figure D.1 Histograms of Level 1 and Level 2 residuals for Exam Scores (pExam)

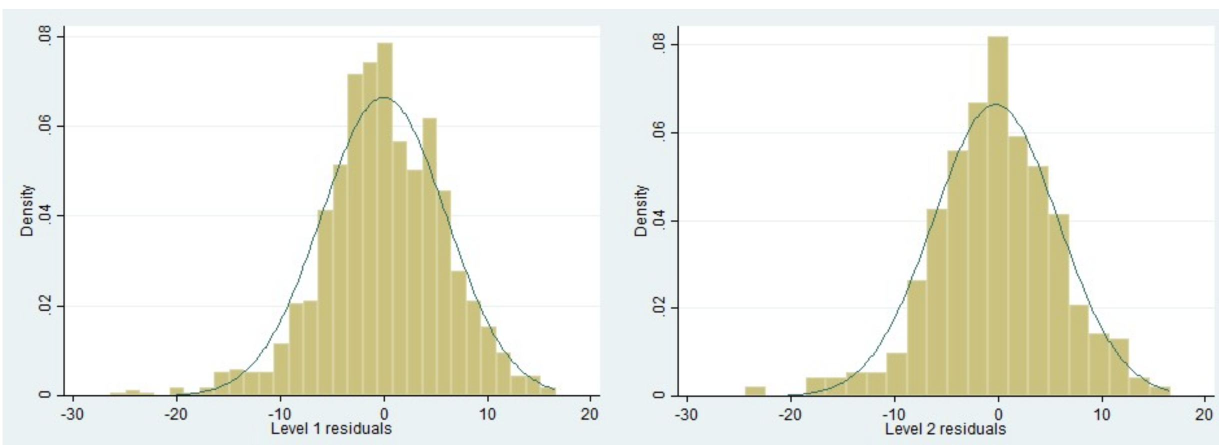


Figure D.2 Histograms of Level-1 and Level-2 residuals for PTP Scores (PTP_Score)

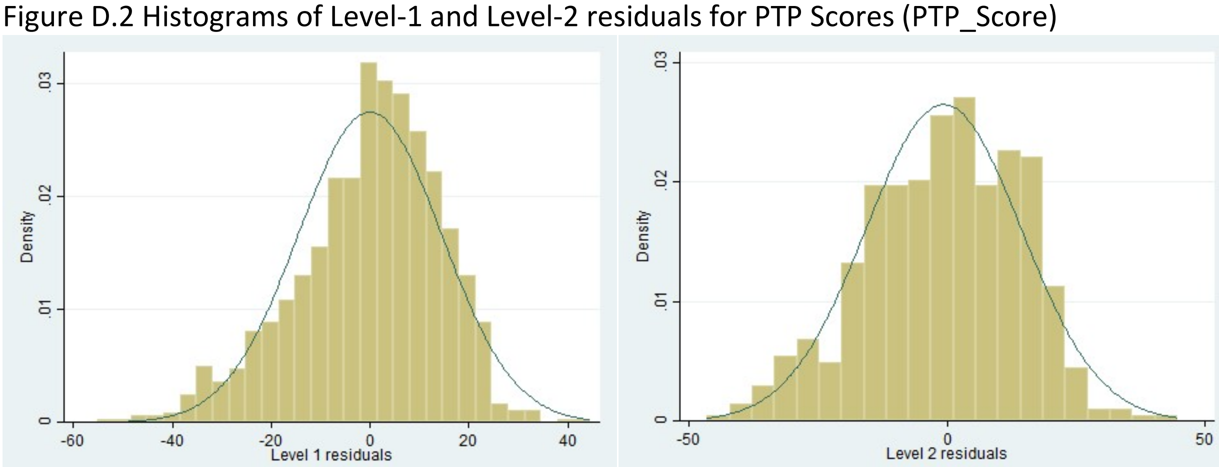


Figure D.3 Histograms of Level-1 and Level-2 residuals for Mechanism Scores

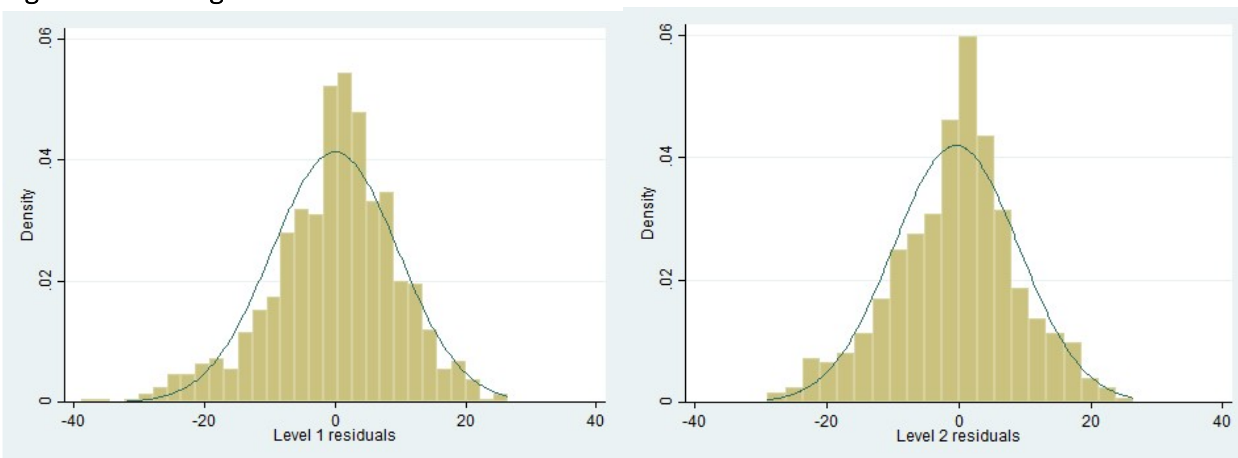


Figure D.4 Histogram of Studentized Deleted Residuals for PTP Q1 – Problem Type. Normality assumption appears to be violated.

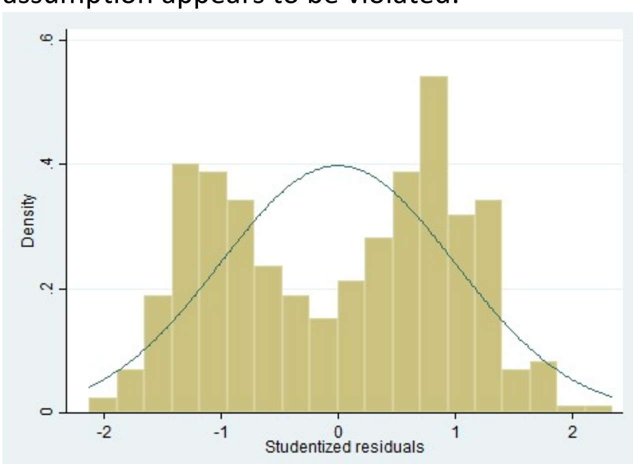


Figure D.4 Histogram of Studentized Deleted Residuals for PTP Q2 – Exercise Type. Normality assumption appears to be violated.

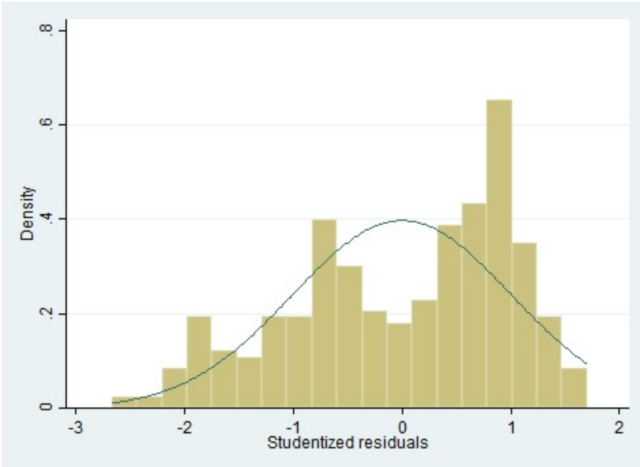


Figure D.5 Scatter plot of PTP Q1 – Problem Type studentized deleted residuals and fitted values. Constant various assumption appears to be violated so robust standard errors used.

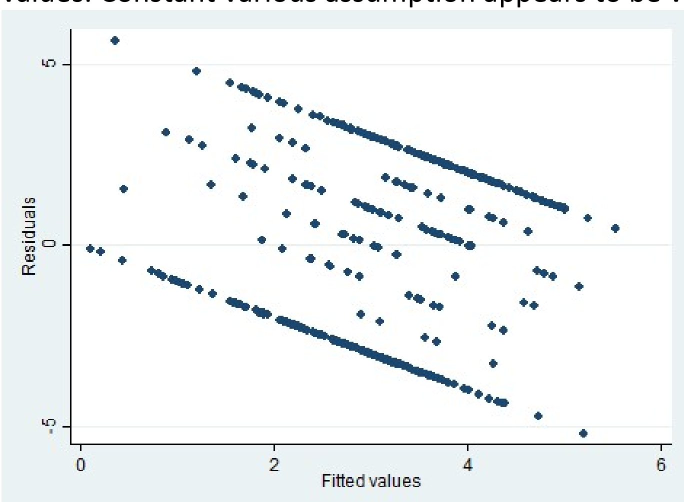
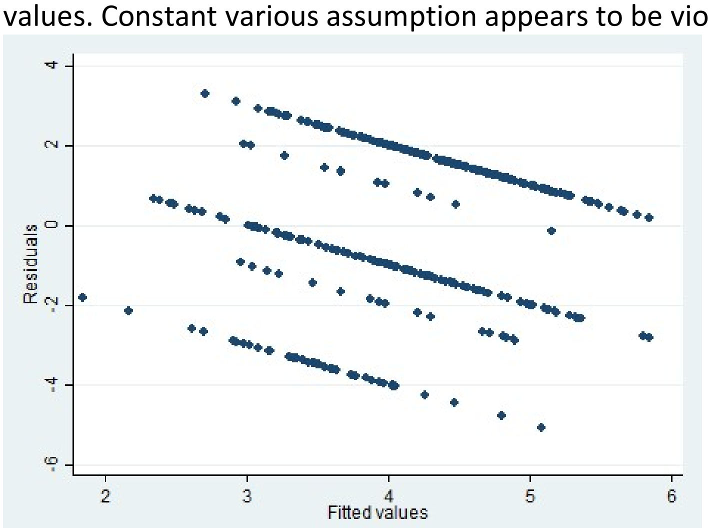


Figure D.6 Scatter plot of PTP Q2 – Exercise Type studentized deleted residuals and fitted values. Constant various assumption appears to be violated so robust standard errors used.

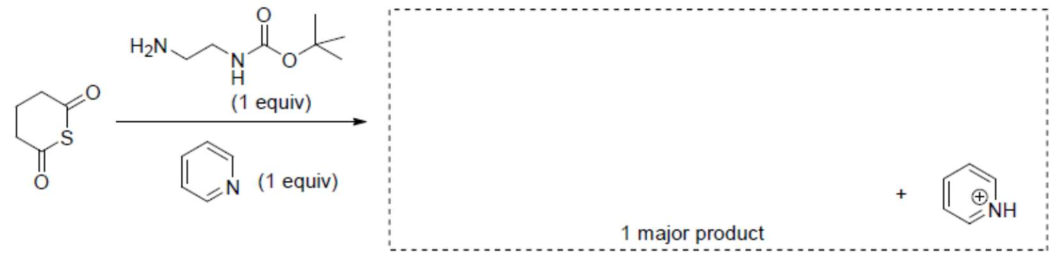


Appendix E: Questions included in Predict-the-Product and Mechanism Scores

Predict the Product Score

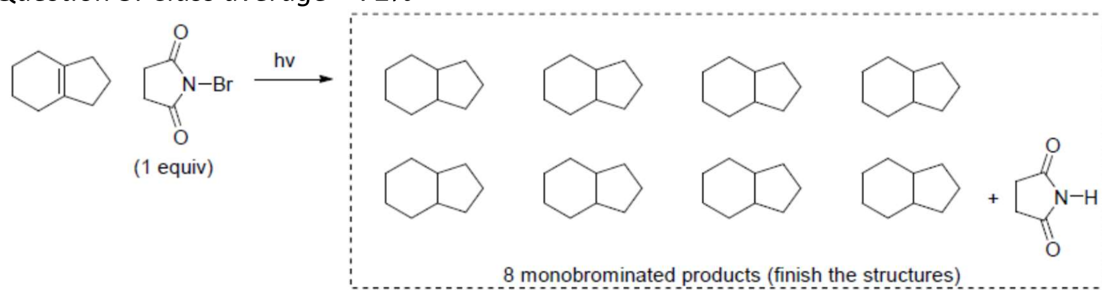
Midterm 1 (Occasion 1):

Question 1: Class average – 47%



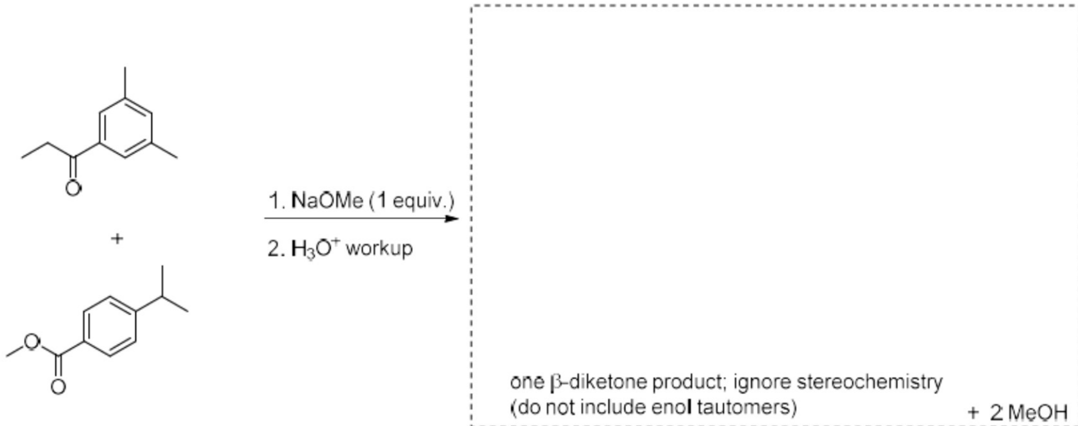
Question 2: Class average – 66%

Question 3: Class average – 72%

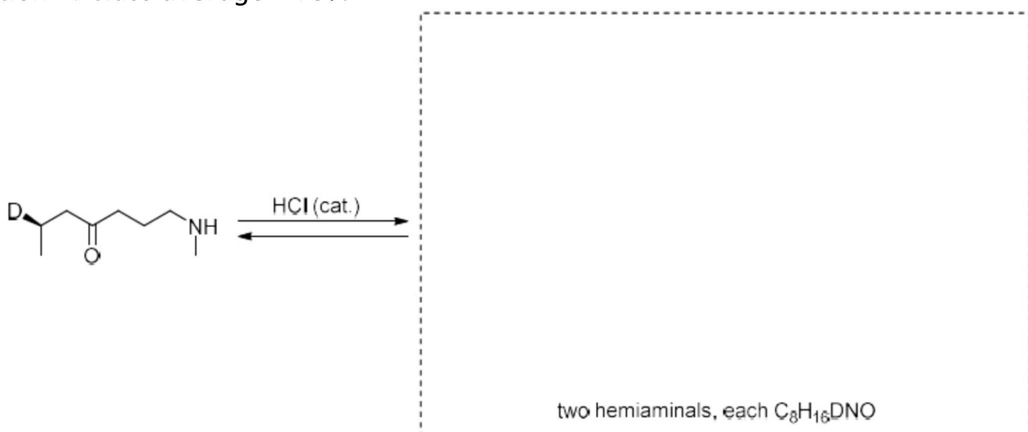


Midterm 2 (Occasion 2):

Question 1: Class average – 83%



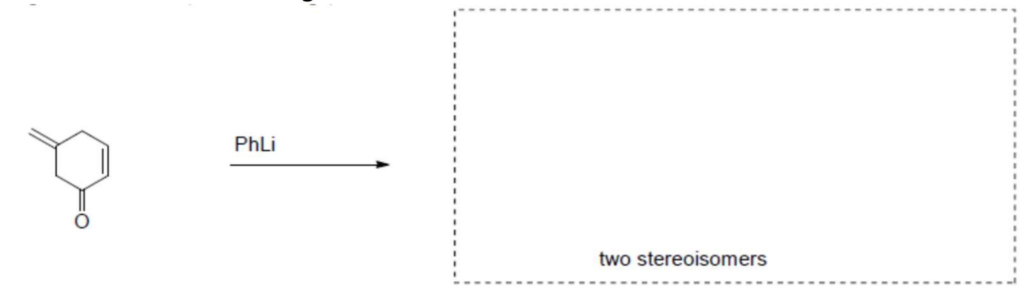
Question 2: Class average – 79%



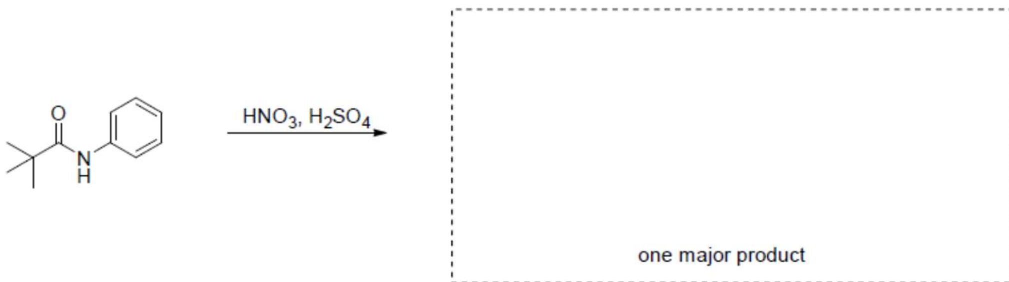
Question 3: Class average – 87%

Final Exam (Occasion 3):

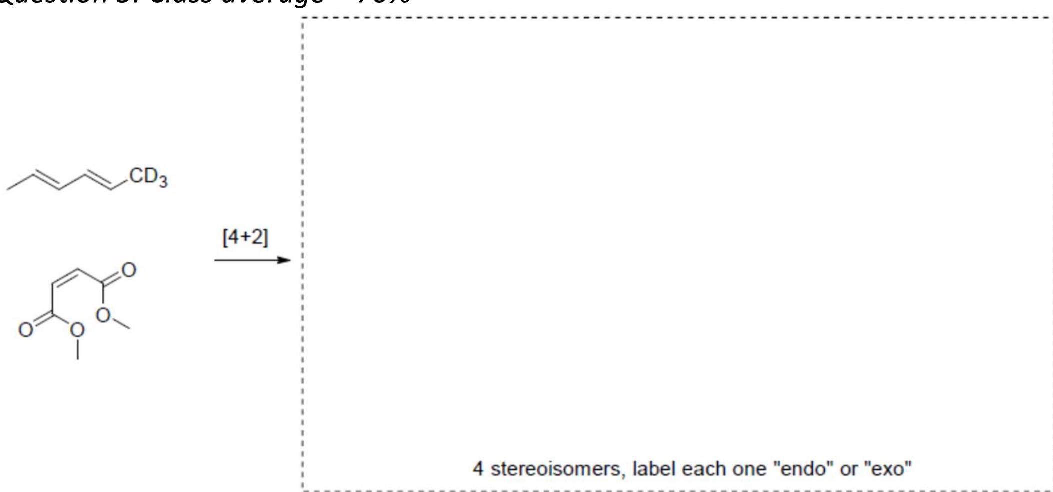
Question 1: Class average – 79%



Question 2: Class average – 68%

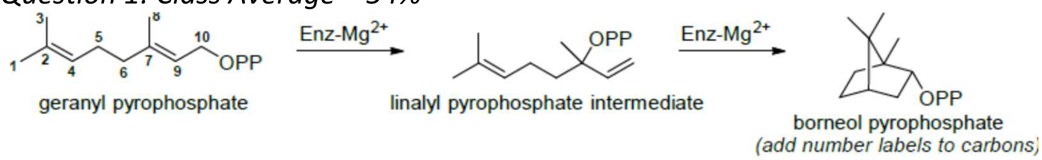


Question 3: Class average – 70%

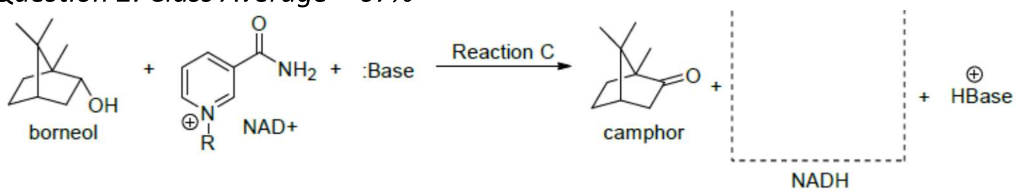


Mechanism Score
Midterm 1 (Occasion 1):

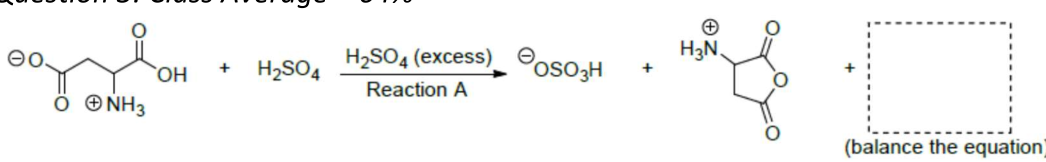
Question 1: Class Average – 54%



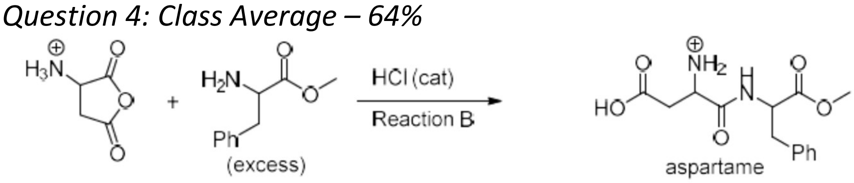
Question 2: Class Average – 67%



Question 3: Class Average – 64%

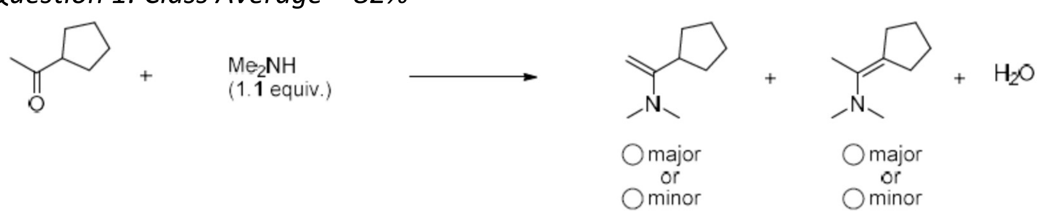


Question 4: Class Average – 64%

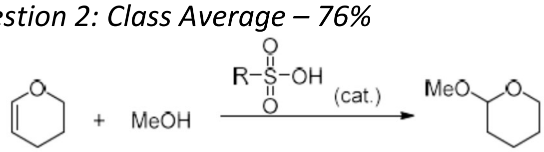


Midterm 2 (Occasion 2):

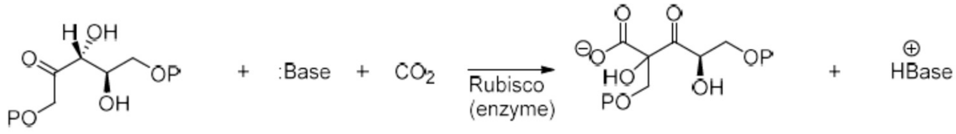
Question 1: Class Average – 82%



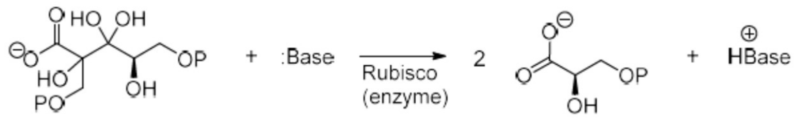
Question 2: Class Average – 76%



Question 3: Class Average – 79%

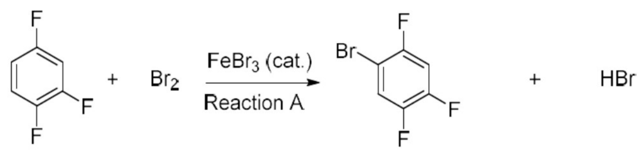


Question 4: Class Average – 82%

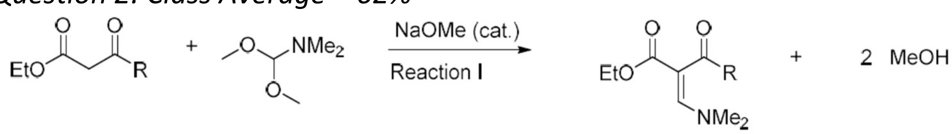


Final Exam (Occasion 3):

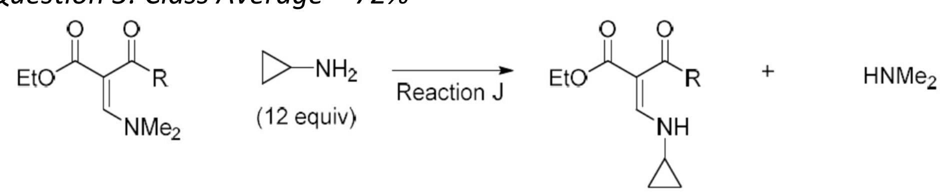
Question 1: Class Average – 80%



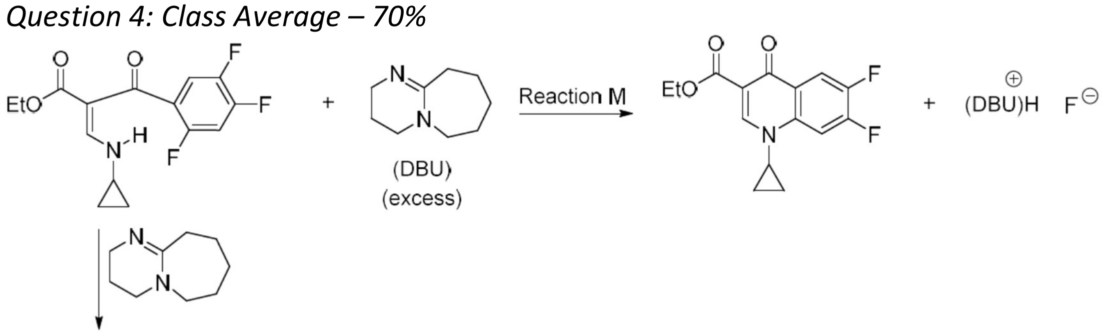
Question 2: Class Average – 62%



Question 3: Class Average – 72%

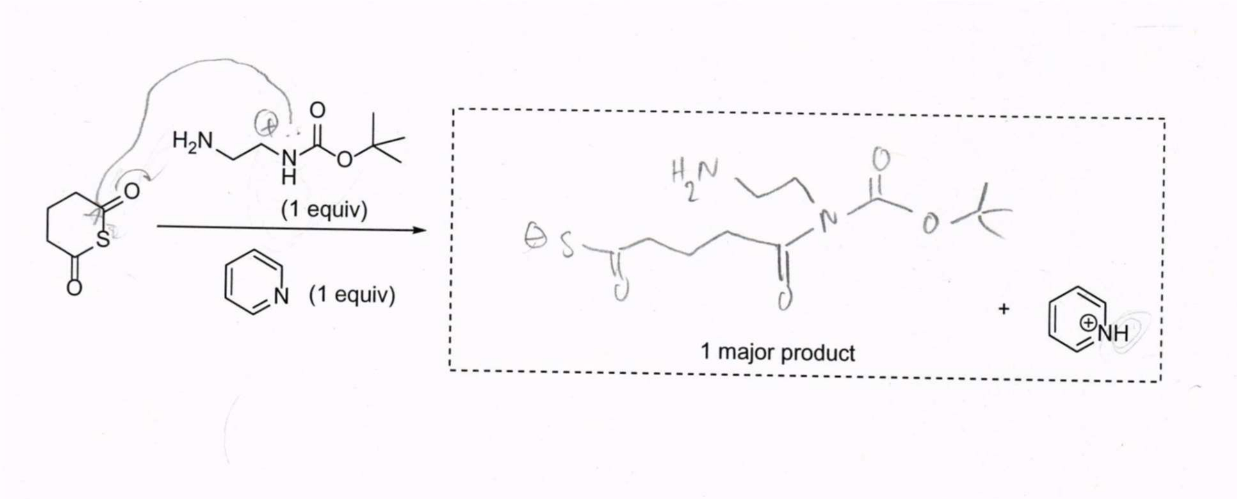


Question 4: Class Average – 70%



Appendix F: Transcripts and Activity Log

Example Problem Transcript



Interviewer: Cool All right, awesome. Uh, you ready to try one more?
Tessa: Okay.
Interviewer: You're doing great.
Tessa: Ugh I'm failing. Okay. Oh, man. Okay. On this question, I remember that I-- there was another question in the exam that was similar to this, so I kind of just copied that part um and I got it right in the end [laugh]. So cause I know the general idea of what this is supposed be but like when this ring is broken up, uh, I wasn't sure exactly how it should look, So I flipped to the other page in the exam and just copied.
Interviewer: Interesting cool um
Tessa: So, in this, um, uh the-- nitrogen over here [points to nitrogen in amide bond and begins to draw an arrow from it to the carbonyl] the loan pair attacks this carbonyl. Uh, oxygen comes up [draws arrow indicating breaking of C-O double bond]. Uh, or the-- the double bond comes up and it's O-- yeah O minus. Uh then the minus comes back down [points to top carbonyl] and I think it-- yeah it kicks out this sulfur [draws breaking of C-S double-bond]. Uh, and you see that there's an H here [points to protonated pyridine]. So that means that originally here [points to amide N] this would be-- this would become like a plus. And I think you would take the hydrogen away here [points to pyridine and then NH] and then leaving you with minus S [begins drawing product] um O, and dun dun and-- and like whatever this is [laugh] Um NH2 uh um. Sorry this is probably not super accurate, but
Interviewer: That's alright
Tessa: Yeah. I think
Interviewer: Okay
Tessa: Yeah. And this gives, like—[points to protonated pyridine] I feel like this gave me a clue on where to deprotonate.
Interviewer: Okay um, So how did you decide which um like nitrogen to attack with?

Tessa: I remember that-- Well, this is the amide one [circles amide bond]. So usually you do that. Um, let's see (2.0) And I guess (1.0) hmm (2.0) I'm not sure if like a resonance argument might apply to it, which would make the-- that nitrogen. Well then it mm be more mm. [laugh] I just remember that you do the-- the amide one.

Interviewer: Fair. Um cool, so what did you do that helped you prepare for this problem?

Tessa: Uh, there were many problems that had this [points to amide bond] not necessarily this pyrimidine down-- I think this is pyrimidine down here [points to pyridine]. Uh, but, uh without even like looking at this part of it [covers pyridine and protonated pyridine], then, like, I knew, like [traces arrow from amide to carbonyl with finger], this is like one of those sort of carboxylic acid derivate-- peptide kind of stuff. So, um-- and based on that [laugh], uh, yeah I just followed the steps. And then, uh, usually I guess they have something else to deprotonate [points to pyridine] but in this case it's pyrimidine and then so you just follow the same steps for a different reaction.

Interviewer: Awesome um what could you have done, if anything, to better prepare for this problem?

Tessa: Um probably do more of the ring practice [points to thioanhydride] I guess so that I feel, like usually rings are like, "oh no. This is here." And so um yeah.

Interviewer: Okay. Do you find like problems with rings like particularly challenging?

Tessa: Uh it yeah like. I should probably number my carbons, and it would make it much easier, um and I know that but I always just kind of, like, trace it with my finger, [begins to trace starting material and product with fingers simultaneously] and I'm like, "Okay, one, two, three, four, Okay, I'm here." and then I just add on the oxygen or something.

Interviewer: Okay that's fair. So is like keeping track of carbons like tricky with rings or?

Tessa: Yeah or more so

Interviewer: Okay

Tessa: Well, yeah, than like regular alkanes but.

Example Activity Log

Person	Decision/Tension Points	Outcome of Decision/Tension Point	Reasoning Used in Decision/Tension Point	Representative Quote/Commentary
Tessa		BP: 3 Predictive (1) Familiarity (2)	Decision: 5 Predictive (2) Familiarity (3)	Tension point: 1 Familiarity vs Chemical - Familiarity
	BP1: On the exam there was a similar problem and she copied the mechanism		Similarity (reaction)	"On this questions, I remember that I—there was another question in the exam that was similar to this, so I just copied that part um and I got it right in the end [laugh]."
	BP2: Understands general idea of mechanism but not in a ring opening context		Recognition (reaction)	"I know the general idea of what this supposed to be but like when the ring is broken up, uh, I wasn't sure exactly how it should look."
	TP1: React with amide or amine		Similarity Chemical – E/N	Order of this step unclear, TP was prompted by follow-up question I: Okay um, so how did you decided which um like nitrogen to attack with? Tessa: I remember that – well this is the amide one [circles amide bond]. So actually you do that. Um, let's see (2.0) and I guess (1.0) hmm (2.0) I'm not sure if like a resonance argument might apply to it, which would the—that nitrogen. Well then it mm be more mm. [laugh] I just remember that you do the—the amide one.
	DP1: React with amide	Decides to react with amide because you usually do that.	Similarity	*Points her in the wrong direction (amine more reactive)* "So, in this, um, uh the—nitrogen over here [points to nitrogen in amide bond]"
	DP2: What to do with amide	Amide N attacks carbonyl and form alkoxide	Similarity (reaction)	*Right direction* "Uh, there were many problems that had this [points to amide bond] not necessarily this pyrimidine down-- I think this is pyrimidine down here [points to pyridine]. Uh, but, uh without even like looking at this part of it [covers pyridine and protonated pyridine], then, like, I knew, like [traces arrow from amide to carbonyl with finger], this is like one of those sort of carboxylic acid derivate--peptide kind of stuff. So, um-- and based on that [laugh], uh, yeah I just followed the steps."
	DP3: What happens with tetrahedral intermediate	Carbonyl reforms and S leaves	Recognition (reaction)	*Right direction* "Uh then the minus comes back down [points to top carbonyl] and I think it—yeah it kicks out this sulfur [draws breaking of C-S bond]
	BP3: Notices protonated pyridine in product box		Predictive - hint	"Uh I see than there's an H here."
	DP4: Charge on amide	Amide N which attacked carbonyl would have a positive charge	Predictive – hint	*Right direction* "So that means that originally here [points to amide N] this would be—this would become a plus."
	DP5: Amide N is deprotonated by pyridine		Predictive- hint	*Right direction* "And I think you would take the hydrogen away here [points to pyridine and then NH] and then leaving you with minus S

Appendix G – Tables of Student Decisions and Ideas in Think-Aloud Interviews

Table G.1: Question 1 Decisions (*n*=53)

	Total (Count)
SO2 will not be produced by the reaction	1
Reaction should be balanced	1
Combine thioanhydride and amide species	1
Form one product	1
Removal of BOC group overall goal	1
BOC group unreactive	2
No reaction at BOC protected amide carbonyl	1
Carbonyl on amide reacts with S	1
Carbonyl is electrophile	1
Sulfur is leaving group	1
Amide N is unreactive	4
Amide N is reactive	2
Amine N is reactive	5
Amide species reacts first	1
Pyridine reacts first	2
Pyridine is a base	6
Pyridine reacts with amide species	3
Pyridine reacts with hydrogens on Ns not Cs	1
Pyridine deprotonates amine N	5
Pyridine deprotonates amide N	4
Pyridine does not deprotonate amine N	1
Deprotonated amide N will have a positive charge	2
Pyridine deprotonates amide N to drive reaction between O and S	1
O needs a driving force to react with S	1
Amine N attacks carbonyl and forms tetrahedral intermediate	8
Amide N attacks carbonyl and form tetrahedral intermediate	4
Amine will not attack carbonyl	1
Carbonyls are equivalent	3
Amide N has a positive charge after reacting with carbonyl	1
Positive and negative charges can coexist	1
Amide N attacks S	1
Amide N did not react with S	1
Alkoxide reacts with protonated pyridine	1
Carbonyl reforms and S leaves	9
Carbonyl reforms and C leaves	1
Pyridine deprotonates activated carbonyl	1
Thioanhydride species ring should break	2

Sulfur in ring is protonated	2
Do not protonate carbonyl	1
Amine is deprotonated by S	1
Ring breaks through formation of a S=C bond and C=O bond breaking	1
Protonated S leaves creating a positively charged carbonyl	1
S- is protonated	1
S- reacts with protonated pyridine	2
S- does not react further	3
Carbanion deprotonates amine	2
Product must be negative	3
Seems correct because there are not stereocenters	1
There won't be a subsequent reaction with the amide species	1
BOC group is removed and forms CO2 and tertiary carbocation	1
Pyridine reacts with tertiary carbocation to form alkene	2
Takes a break	2
Gives up	3

Table G.2: Question 1 Other Ideas (*n*=38) – Brought up during brainstorming and tension points

	Total (Count)
Mechanism for BOC removal	1
Lack of driving force for BOC removal	1
BOC group present	2
Recognizes sulfur	1
Thioanhydride looks like an anhydride	1
Where does pyridine take H from?	1
Resembles an amino acid reaction but no OH	2
Thioanhydride resembles SO2	2
Amide species can react at amine end	1
Sulfur reacts with oxygens	1
Amide looks like an amino acid	1
Identified amide bond was formed	1
Amide bonds are stable	1
Amide bond has a on oxygen next to it so something could attack	1
One major product	2
Pyridine protonated in product	4
Amide species reacts at one of two N's	1
Unsure where it will react on the anhydride	1
Unsure if ring opens	1
S can have a minus charge but doesn't want to	1

Oxygen protected by BOC group	2
Similar problem on the exam and she copied the mechanism	1
Understands general mechanism but not in a ring-opening context	1
Negative carbon is unusual	1
Looks at reagents and starting material	1
BOC group intimidating	1
S- feels wrong	1
Pyridine has to deprotonate something	1
Sulfur's a withdrawing group	1
Deprotonated amine is a strong base	1
Draws loan pairs on oxygens and sulfur	1
Thioanhydride resembles succinimide	1
Figure out where H comes from and push two remaining things together	1
Pyridine will be protonated	1
Could deprotonate amine with carbonyl	1
Sulfur could attack carbonyl	1
How does S react?	1
Formation of =N bond and OH leaving	1

Table G.3: Question 2 Decisions (*n*=34)

	Total
Four products	1
Stereochemistry present	1
Use reagents in order	1
Li won't act as organolithiate	1
Li reacts with Br	5
Li reacts with Br to form carbanion	6
Li reacts with Br and it's equivalent to carbanion	2
Li reacts with Br and leaves	1
Something needs to replace Br	1
LiBr leaves and forms carbanion	4
LiBr leaves and forms carbocation	1
Li doesn't react with another site	1
Li replaces Br	1
Br leaves and forms a carbocation	2
Positive charge on Li	1
Grignard does not attack twice	1
Resonance leads to two reactive sites	11
Carbanion reacts with ketone	1
Carbanion reacts with ketone and forms tetrahedral intermediate	12
Carbocation reacts with ketone and forms tetrahedral intermediate	1
Intermediate reacts with ketone and forms tetrahedral intermediate	1
Four structures means stereocenters	5
Two structures come from each resonance site	1
Different versions (wedged and dashed) bonds forms	4
Stereochemistry creates wedged and dashed products	4
Alkoxide carbon is not a stereocenter	4
Carbon at ring is a stereocenter which leads to two products	5
Change orientation of alkene side-chain to get more products	3
Orientation of alkene side-chain doesn't change	4
No E1 or E2 reactions occurring	1
No rearrangement of double bonds	1
No identical products	1
H3O+ protonates alkoxide	13
Gives up	2

Table G.4: Question 2 Other Ideas (n=26) – Brought up during brainstorming and tension points

	Total
Four structures is confusing/stressful	2
Only one product but hint says four structures	2
Always forget resonance but luckily I didn't	1
Sequence of steps feels different	1
Li is in excess so it might with other sites	1
Three steps	1
Recognized acetone	1
H3O+ protonates at the end and forms OH	1
Uncertain what Li does because it's usually seen in LiAlH4 or there's an Mg	1
Acetone adds to Li	1
Need something to help Br leave	1
Question from Organic Chemistry 1	1
Resonance is the Organic Chemistry 2 step	1
Do grignards behave like carbanions and have resonance	1
Four structures	2
Organolithiates react with H's from allylic bonds	1
Li usually used to make something negative	1
Where would carbocation react	1
Four structures means different isomers	1
Reagents not normally seen together	1
Li helps Br leave	1
Rarely Sn2 so when there's a leaving group it's not concerted	1
Four structures means resonance, another reactive site, or E1 or SN1, or carbocation rearrangement	1
Dislikes wedges and dashes and ignores them	1
With multiple structures can make diastereomers and enantiomers	1
Some sites shouldn't have their stereochemistry change	1