University of South Florida
Scholar Commons

Graduate Theses and Dissertations

Graduate School

3-1-2020

Asking *Why*: Analyzing Students' Explanations of Organic Chemistry Reaction Mechanisms using Lexical Analysis and Predictive Logistic Regression Models

Amber J. Dood
University of South Florida

Follow this and additional works at: https://scholarcommons.usf.edu/etd

Part of the Chemistry Commons, and the Education Commons

Scholar Commons Citation

Dood, Amber J., "Asking *Why*: Analyzing Students' Explanations of Organic Chemistry Reaction Mechanisms using Lexical Analysis and Predictive Logistic Regression Models" (2020). *Graduate Theses and Dissertations.*
https://scholarcommons.usf.edu/etd/8189

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.

المنارة للاستشارات

www.manaraa.com

Asking *Why*: Analyzing Students' Explanations of Organic Chemistry Reaction Mechanisms using Lexical Analysis and Predictive Logistic Regression Models

by

Amber J. Dood

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosopy
Department of Chemistry
College of Arts and Sciences
University of South Florida

Major Professor: Jeffrey R. Raker, Ph.D.
Wayne C. Guida, Ph.D.
Jennifer E. Lewis, Ph.D.
Luanna B. Prevost, Ph.D.

Date of Approval:
February 26, 2020

Keywords: tutorial design, constructed-response, Lewis acid−base, unimolecular substitution, education research

© Copyright 2020, Amber J. Dood

المنارة للاستشارات
www.manaraa.com

Dedication

"There's a million things I haven't done, but just you wait."

-Alexander Hamilton (in *Hamilton: An American Musical*)

This work is dedicated to everyone who helped me along the way. I would not have gotten here without the immense support I have received from those who care about me. Specifically, I want to dedicate this dissertation to my dad, who I know would be proud of how much I have overcome in order to get here.

www.manaraa.com

Acknowledgments

There are many people without whom this work would not have been possible. So many family, friends, and colleagues have supported me before and during graduate school. I would like to thank each every person who has ever encouraged me to continue to pursue my goals, as there have been many instances in my life where I wanted to simply give up.

First and foremost, my parents. Both of my parents encouraged me to work hard and achieve my goals, even when those goals were not necessarily well-defined. My dad supported me in every possible way while he was still able to do so and for that I am eternally grateful. It never made a difference to him that I was a girl. Even when my goals included winning motocross championships, he supported me wholeheartedly. My mom, also, has always supported me and continues to support me on a daily basis. She deserves much more gratitude than I tend to provide. Thank you also to my sister, Anna, my oldest friend, who for some reason continues to put up with me.

I want to extend special thanks to my husband, John Dood, for moving over 1,000 miles with me so I could attend USF and putting up with me the past three years. I am forever grateful for the meals cooked, laundry cleaned, groceries purchased, and everything else I lacked the time to do while in school. Spending the first three years of our marriage with me in graduate school was tough, but we did it. Additionally, the work presented here could not have been completed in its present form without the immense help of John's coding skills and advice. The Python program presented here and used to develop my predictive models would not exist without John. John, thank you for your neverending support and love in everything I do.

المنارة للاستشارات

www.manaraa.com

Thank you to my mentors in my previous chemistry research experiences who encouraged me to go on to graduate school. Specifically: Jason Gillmore and Brent Krueger during my time as a student at Hope College, Tom Guarr and Laura Ives while I worked at the Michigan State Bioeconomy Institute, and Justin Shorb who guided me through my first Chemistry Education Research experience while I was teaching at Hope. I would never have had the guts to take on graduate school without your support (and letters of recommendation!).

Thank you to my CER colleagues were invaluable during my time at USF for advice, feedback, and friendship. Without you, my time in graduate school would have been more challenging and much more boring! I appreciate all of the support the USF CER community has provided to me throughout my graduate school career.

The students who participated in the many surveys required to collect data for this dissertation also deserve my thanks, along with Dr. Kimberly Fields and Dr. Daniel Cruz-Ramírez de Arellano, who allowed me to collect data in their courses. Without their constant support as co-authors, this dissertation could not have happened. I also need to thank them for allowing me to be a teaching assistant in their courses. I learned so much about teaching organic chemistry by working with them.

I would also like to thank my committee, who supported me throughout my graduate career. Dr. Lewis, thank you for always pushing me a little bit further than I was comfortable going during committee meetings. Dr. Guida, thank you for being so genuinely interested in my work and for encouraging me to think about real applications in the classroom. Dr. Prevost, thank you for your guidance in lexical analysis techniques and helpful feedback during committee meetings. I would also like to thank Dr. Prevost's former student, Dr. Kelli Carter, who was kind enough to work with me on learning to use SPSS Modeler and troubleshooting the many issues I experienced with it.

Finally, I would like to thank my major professor, Jeff Raker, without whose encouragement and expertise this never could have happened. They say that choosing your advisor is the most important decision

المنارة للاستشارات

www.manaraa.com

you will make in graduate school and I wholeheartedly agree. My decision to to work with Jeff was one that I would make again every time. I am forever grateful for that he gave me the opportunity to work with him. Jeff, thank you for always supporting and advocating for me while also being stern with me when I needed it. Thank you for letting me cry in your office and keeping up with and always supporting my ever-changing career goals. Thank you for encouraging me to do things other than work, even if I ignored you most of the time. And finally, thank you for helping me become a better researcher and person through my graduate school experience. I will truly miss your guidance, reassurance, and excessive semi-colon use.

للاستشارات المنارة

www.manaraa.com

Table of Contents

المنارة للاستشارات

www.manaraa.com

المنارة للاستشارات
www.manaraa.com

www.manaraa.com

List of Tables

www.manaraa.com

List of Figures

المنارة للاستشارات

www.manaraa.com

Abstract

In order to evaluate student understanding of chemical reactions and reaction mechanisms, we must ask students to construct explanations of mechanistic representations. Grading written assessments is time-consuming, which limits their use in the classroom. Lexical analysis and logistic regression can be used to develop models that predict human scoring for constructed-response items. In this work, students' responses to constructed-response items about what is happening and why in two types of organic chemistry reaction mechanisms are explored: acid−base proton-transfer (Chapter 3) and a unimolecular substitution (i.e., S_N1; Chapter 5). The acid−base proton-transfer item was scored for use of the Lewis acid−base model. The S_N1 item was scored for three levels of explanation sophistication.

The utility of predictive text analysis models for development of instructional materials is exemplified in Chapter 4. A research-based tutorial was designed to increase student use of the Lewis acid−base model in written explanations. The predictive model was employed to efficiently analyze students' responses before and after the tutorial to assess the tutorial effectiveness, finding a positive impact on use of the Lewis model.

The lexical analysis and logistic regression techniques used in this dissertation can be applied to many other contexts to produce predictive models which can be used in the classroom to help instructors and students evaluate the quality of their explanations. Instructional tools like the Lewis acid−base tutorial in Chapter 4 can be used to help students construct the knowledge necessary to explain reaction mechanisms. Many have called for the use of writing in the science classroom; techniques like those presented in this dissertation pave the way for open-access, efficient tools to provide feedback for writing prompts in organic chemistry courses and classes in the sciences as whole.

المنارة للاستشارات
www.manaraa.com

Chapter 1

Introduction

It is well-documented that students in traditional organic chemistry classes experience difficulty in understanding reaction mechanisms (Bhattacharyya, 2014; Bhattacharyya and Bodner, 2005; Bodé *et al.*, 2019; Caspari *et al.*, 2018a, 2018b; Flynn and Featherstone, 2017; Grove *et al.*, 2012; Kraft *et al.*, 2010; Strickland *et al.*, 2010). The work in this dissertation explores how students describe and explain acid−base (Chapter 3) and unimolecular substitution reactions (Chapter 5). Additionally, the work introduces predictive scoring models which use lexical analysis and logistic regression techniques to analyze and score students' responses based on coding schemes determined through qualitative analysis of the data sets. The work paves the way for automated scoring of students' written responses in organic chemistry courses and has applications in formative assessment. Automated scoring models can also be used to tailor educational activities based on the content of responses, work that is explored in Chapter 4.

Organic chemistry mechanisms, or the stepwise movement of electrons through the process of the reaction, have been taught to organic chemistry students for decades to explain the how and why behind the many reactions students learn in the course (Kermack and Robinson, 1922; Morrison and Boyd, 1959; Wheeler and Wheeler, 1982). Though instructors see these mechanisms, portrayed through the electron pushing formalism (EPF), as a way to connect all of the reactions in the course together in a coherent way occur (Bhattacharyya, 2013; Cooper *et al.*, 2016), chemistry education research (CER) has shown that students do not understand the meaning behind the arrows (e.g., Bhattacharyya and Bodner, 2005). Instead, students understand the arrows to be another part of the reaction to memorize, rather than seeing how the

www.manaraa.com

EPF shows the movement of electrons (e.g., Grove *et al*., 2012). Students have been found to memorize the sequence of intermediates between reactant and product, draw out these intermediates, and then add in the arrows afterward as if "decorating" the structures with arrows (Grove *et al.*, 2012). Organic chemists use the EPF as a tool to show and predict reactivity, using the arrows to show the electron movement that occurs to go from the reactant to each intermediate and then finally to the product (e.g., Kozma, 2003).

This creates a disconnect between instructors' expectations and what students actually understand. Students are able to reproduce pictorial representations of organic chemistry reaction mechanisms correctly from memory (Grove *et al.*, 2012). Thus, when instructors assess students on what instructors might think is understanding of mechanisms (i.e., asking students to "provide the mechanism" for a reaction), students can come up with the correct answer without understanding what the arrows mean. In order to assess whether or not students understand the meaning behind the arrows, students need to be asked to explain (e.g., Bodé *et al.*, 2019; Caspari *et al.*, 2018a; Crandell *et al.*, 2018).

Studies about student explanations of organic chemistry reaction mechanisms have appeared in the literature based on verbal descriptions of mechanisms through interviews (e.g., Bhattacharyya, 2014; Bhattacharyya and Bodner, 2005; Caspari *et al.*, 2018a, 2018b; Cruz-Ramírez de Arellano and Towns, 2014; Webber and Flynn, 2018; Weinrich and Sevian, 2017) and analysis of examination questions (Bodé *et al.*, 2019; Cooper *et al.*, 2016; Crandell *et al.*, 2018; Grove *et al.*, 2012; Webber and Flynn, 2018). In general, these studies have shown that many students do not understand the meaning behind the EPF and simply memorize the arrows along with the reactions.

1.1 Revised curricula

Curriculum redesigns have placed a greater emphasis on the explanations behind reactions in organic chemistry. The "Mechanisms Before Reactions" curriculum (Flynn and Ogilvie, 2015) teaches mechanisms

المنارة للاستشارات
www.manaraa.com

and the EPF before specific reactions are taught, encouraging students to understand how and why electrons flow as they do in reactions before they begin memorizing the large amount of reactions presented to them in organic chemistry class. The order of topics in the course is also changed from that of typical organic chemistry courses, with reactions grouped by their mechanism rather than by functional group transformation. The hope is that students will begin to see the patterns between reactions and therefore be able to achieve deeper learning. The Organic Chemistry, Life, the Universe, and Everything (OCLUE) curriculum is another example of a total redesign of the sophomore organic chemistry course (Cooper *et al.*, 2019). A continuation of the General Chemistry, Life, the Universe, and Everything (CLUE) course, OCLUE focuses on three-dimensional learning and emphasizes biologically important mechanisms. The course seeks to engage students in scientific practices as defined by the *Framework for K-12 Science Education* (National Research Council, 2011) and asks students to construct models and explanations with a strong focus on causal mechanistic reasoning.

The authors of the Mechanisms Before Reactions and CLUE/OCLUE curriculums have asked students to explain reaction mechanisms in research studies, finding encouraging results related to the explanation ability of students who participated in the revised curricula. Galloway *et al.* (2017) found that students in the Mechanisms Before Reactions curriculum scored higher on both familiar and unfamiliar mechanism questions than those who were not in the revised curriculum, although the authors could not conclude that the increase in success was directly related to the change in curriculum. Crandell *et al.* (2018) found that students who took the general chemistry CLUE curriculum retained their ability to use causal mechanistic reasoning when describing mechanisms and were able to correctly produce mechanistic arrows for acid−base proton-transfer reactions between HCl and H_2O and NH_3 and BF_3 during the summer between completing general chemistry 2 and beginning organic chemistry 1. Students who had participated in the general chemistry CLUE curriculum were more likely than students who had taken other general chemistry courses to employ causal mechanistic reasoning and provide correct mechanistic reasoning when describing

المنارة للاستشارات
www.manaraa.com

what happens and why in an acid−base proton transfer reaction and to provide correct mechanistic arrows at the beginning of organic chemistry 1. CLUE students continued to outperform non-CLUE students on these tasks at the end of organic chemistry 2 even though both cohorts participated in the same, non-revised organic chemistry curriculum. To date, there has been no published work related to the performance of OCLUE students as compared to traditional organic chemistry students, but the causal reasoning ability of general chemistry CLUE students is promising. The enhanced mechanism-solving ability of both CLUE and Mechanisms Before Reactions students is encouraging for the success of revised organic chemistry curricula that focus on causal explanations rather over traditional organic chemistry curricula, which may lead students to memorization over deep understanding.

Instructors must make it known to students that understanding the explations behind reaction mechanisms is a vital part of understanding organic chemistry. As we know that students can memorize reaction mechanisms and regurgitate the mechanisms without understanding the meaning behind the structures and arrows, it is important to specifically ask students *why* reactions are occurring. While asking students to write is one way to do this, instructors who teach large courses may find it impractical to read student writing on examinations or other assignments. One way to analyze student writing quickly is through automated text analysis.

1.2 Computer-assisted scoring of written responses

Computers can be trained to look for certain keywords and text patterns in writing (Popping, 2000). Additionally, computers can be trained to assign categories and overall scores to written work using lexical analysis and logistic regression techniques (e.g., Prevost *et al.*, 2016). There are many constructed-response items in the Automated Analysis of Constructed Response (AACR) database that ask students to write about biology, biochemistry, chemistry, and statistics topics (beyondmultiplechoice.org). The database is mainly

المنارة للاستشارات
www.manaraa.com

biology-focused, with non-biology related items being highly relevant to topics discussed in introductory biology. Items specific to organic chemistry are not currently included in the database. Several published works discuss the development of specific items included in the database and the development of predictive models for the items (Ha *et al.*, 2011; Haudek *et al.*, 2012, 2011; Kaplan *et al.*, 2014; Prevost *et al.*, 2012, 2013, 2016; Weston *et al.*, 2015).

There are different types of logistic regression predictive models that can be used to score written responses, including binomial, multinomial, and ordinal logistic regression (Liu, 2015). For example, Prevost *et al.* (2016) used multinomial logistic regression to automatically score student responses to three items about how an alteration in a DNA sequence impacts replication, transcription, and translation. Responses were scored as correct (*1*), incomplete (*2*), or incorrect (*3*). In the work presented in Chapter 3, a binomial logistic regression model was used to determine if a student response evoked a Lewis acid−base model (*1*) or not (*0*). Ordinal logistic regression can be used if the variables to be predicted have an order. For example, if the prediction were to score responses as on a Likert scale from 1-4 where *1* is "strongly disagree", *2* is "disagree", *3* is "agree", and *4* is "strongly agree" (Liu, 2015).

Some requirements must be met in order to successfully construct a predictive model. First, a large number of responses must be collected and human-scored in order to amply train and test the model. Though there is no set number of responses needed, the number used in published works tends to be around 1,000 (Ha *et al.*, 2011; Prevost *et al.*, 2016). Typically, about 70% of the responses make up the training set and 30% are left aside to become the testing set. Having around 1,000 responses allows a large amount of responses for the training set while still having a testing set that resembles the class size of many large lecture courses. Second, data sets should have an ample amount of responses for each score in order to properly train the model. Ideally, a data set for a binomial logistic regression model would have half of all responses human-scored as *0* and the other half human-scored as *1*. This important because if only 10% of responses were human-scored as *0* and 90% of responses were human-scored as *1* in a binomial logistic regression, the

المنارة للاستشارات
www.manaraa.com

predictive model could have 90% accuracy as compared to human scoring, but actually just be scoring every response as *1*, having 100% accuracy for responses scored as *1* and 0% accuracy for responses scored as *0*. Some variation from equal numbers of responses across scores may be okay, as long as attention is paid to accuracy levels per score and there are not large differences in accuracy across scores. Third, the model must reach a high level of overall accuracy. While there is no set accuracy threshold, published models typically reach at least 70% accuracy and strive for greater than 80%.

Predictive models require a large amount of time to construct and are specific to the constructed-response item for which they were developed. Although some models can be expanded to include several highly similar items (see Chapter 4 and Chapter 5), most items require their own predictive model. The time to construct the model is a limitation for predictive model use; however, if models are developed with high accuracy, time spent grading can be greatly decreased by using the predictive model to score formative assessments and students can receive feedback in a more timely manner. Models can also be used to develop educational materials, as evidenced by the tutorial presented in Chapter 4. From a research perspective, information about the content of students' responses can be explored at a greater depth by analyzing the categories used in the logistic regression model(s) for their frequency in responses, coefficients, and odds ratios. Due to the utility of lexical analysis and predictive logistic regression models, these techniques are central to the work presented here.

1.3 Overview of the work

This work began with a publication by Cooper *et al.* (2016), which asked general chemistry students in a revised curriculum (see Cooper and Klymkowsky, 2013) to describe what is happening in an acid−base proton-transfer reaction and why it is happening, how they would classify the reaction, and to draw the mechanistic arrows for the reaction. Students described the reaction between hydrochloric acid (HCl) and

المنارة للاستشارات
www.manaraa.com

water (H_2O) in a variety of ways. Some students focused on a Brønsted-Lowry acid−base model while others focused on the Lewis acid−base model. Sophistication of students' answers ranged from simple descriptions (i.e., the what) to mechanistic explanations (i.e., the what and how) and causal explanations (i.e., the what and why). The Cooper *et al.* (2016) study collected data via an online survey system that could accept both written and drawn responses.

As the starting point for my research, the same item (reaction of HCl and H_2O) and similar items which asked about S_N1 and E1 reaction mechanisms were given to students at USF on a series of written homework assignments in a traditional first-semester organic chemistry class. When looking at the student responses about the reaction of HCl and H_2O, the coding scheme from the Cooper *et al.* (2016) study was kept in mind. However, I quickly found that the students' responses in my data set would not fit into the same categories used by Cooper *et al.*, who put responses in mutually exclusive categories based on acid−base model used. Instead, most students were using mixed models. Thus, responses could not be sorted into a Brønsted-Lowry or a Lewis category, because many students were using a combination of the two models. Additionally, some students were using an Arrhenius model, either by itself or in combination with one or more of the other models.

Each response was coded for whether or not it invoked Arrhenius, Brønsted-Lowry, and Lewis acid−base models. For example, a response could be coded as using all three models, two of the models, one of the models, or none of the models. Students took an exam which included acid−base topics shortly after completing the homework assignment which included the constructed-response item about HCl and H_2O. Looking at students' scores on the exam as a whole and specifically their scores on the acid−base related items, students who invoked the Lewis acid−base model in their explanations scored significantly higher than those who did not. This did not have to be sole use of the Lewis model; students who used the Lewis model combined with other models were also included in the group with significantly higher scores.

المنارة للاستشارات
www.manaraa.com

As there was a link between acid−base model use and performance on the exam, it made sense to make this assignment accessible to administer to large classes. Though it had originally been an assignment in a large lecture course, the assignment itself was graded for completion. Analysis of the large amount of responses had taken a lot of time; the time required was impractical for an instructor to spend grading a homework assignment worth a very small percentage of the students' final grades. Looking at the current work being done by the AACR group, lexical analysis seemed like a viable option. As the current data were handwritten, more responses were collected via an online survey to produce electronic responses to use to develop a model. Conveniently, there were already some researchers at USF working with the AACR group. With their assistance, I developed a logistic regression model using SPSS Modeler that could predict whether or not a students' response to the prompt was invoking the Lewis acid−base model or not with high accuracy when compared to human coding. The details of this predictive model, as well as analysis of the homework assignment responses, are presented in Chapter 3.

After the encouraging results of the predictive model, there were two logical directions for the project: (1) use the predictive model to build tools for the classroom and (2) create similar predictive models for items with different reaction mechanism types. Due to the link between Lewis acid−base model usage on the constructed-response item and higher exam scores, a tutorial which increased use of the Lewis acid−base model in responses to the constructed-response item seemed like a fitting next step. Using prior work on student understanding of acids and bases from the literature, a tutorial was developed to move students from non-use of the Lewis acid−base model to use of the Lewis acid−base model when responding to the constructed-response item. To test the utility of the tutorial, students needed to be asked to respond to the constructed-response item, complete the tutorial, and then respond to a similar (i.e., "cloned") constructed-response item. In preparation for the tutorial, additional data were collected: responses to constructed-response items which asked students to explain what happens in an acid−base proton-transfer reaction and why, except this time using hydrobromic acid (HBr) and hydroiodic acid (HI) instead of HCl. Once these

www.manaraa.com

responses were collected, I edited the predictive model so that it could predict Lewis acid−base model usage for responses about the cloned items in conjunction with the original item. The cloned items were then used when developed a tutorial, which is detailed in Chapter 4.

Rather than have certain students take the tutorial and others not take the tutorial (i.e., have a control group), a McNemar test was used to see if there were differences in Lewis acid−base use before and after the tutorial (McNemar, 1947; Sheskin, 2011). The tutorial increased use of the Lewis acid−base model significantly across three different populations (organic chemistry 1 students before summative assessment on acids and bases, organic chemistry 1 students after summative assessment on acids and bases, and organic chemistry 2 students). There was also no significant difference in Lewis acid−base model usage immediately post-tutorial and Lewis acid−base model usage after a three-week time delay. Though there are many other factors that may have contributed to students retaining their increased usage of the Lewis acid−base model (e.g., three additional weeks of instruction in organic chemistry), the work does show that lexical analysis and predictive logistic regression models can be used to develop effective tools for use in the classroom.

To begin developing additional predictive models for similar reaction types, student responses were collected for items very similar to the original acid−base prompt and asked students to explain what happens and why for a unimolecular substitution (S_N1) reaction. Due to the difference in student responses as compared to the responses about the acid−base proton-transfer reaction, a new coding scheme was necessary. I decided to again allow the coding scheme to arise from the data. By printing out all responses, cutting them into strips, reading through the responses repeatedly, and sorting them by hand, a tri-level scoring scheme emerged. Some themes throughout the responses included discussion of the leaving group, carbocation, nucleophile and electrophile, and proton-transfer. Discussion of each of these components came through at levels which fit the overall tri-level scoring scheme. A detailed discussion of the scoring scheme and how students described each component can be found in Chapter 5.

المنارة للاستشارات
www.manaraa.com

Using the program SPSS Modeler, I then began to develop a predictive model for the tri-level scoring scheme. Predictive modeling for three levels generally calls for an ordinal logistic regression model. In this case, two binomial logistic regression models were more accurate. One model predicted Level 1 vs. Level 2/3, and the other model predicted Level 1/2 vs. Level 3. Predicted Level 2 responses were determined by combining the two models. In the process of developing these models, I decided that I wanted more control over exactly what the program was doing when developing the predictive model than SPSS Modeler would allow. Additionally, licenses for SPSS Modeler are pricey and I wanted anyone to be able to use the models I developed with financial constraints. With the assistance of a software engineer, I developed a program to build predictive models using Python. In the new program, I was able to reproduce the two binomial logistic regression models I had created in SPSS Modeler and edit them to achieve even higher accuracy than I had with the acid−base predictive model. Details of the program and the two binomial logistic regression models used to predict levels for the S_N1 reaction can be found in Chapter 5.

With the Python program, other researchers can develop predictive models for constructed-response items of their choice. The models can be used to develop learning tools similar to the Lewis acid−base tutorial described in Chapter 4 or to provide students with immediate feedback about their written responses. This feedback can be as detailed as the instructor or developer of the predictive model desires.

In Chapter 2, I describe in detail the statistical methods used in the studies in this dissertation as well as the details of how the Python program works. Then, my research is presented in three chapters. In Chapter 3, I discuss the analysis of students' explanations of an acid−base proton-transfer reaction mechanism and the development of a model to predict use of the Lewis acid−base model in written explanations. In Chapter 4, I detail work on the development of a tutorial to increase students' use of the Lewis acid−base model when explaining the acid−base proton-transfer reaction. In Chapter 5, students' explanations of S_N1 reactions are discussed, along with a predictive model which scores three levels of explanation sophistication. Chapter

المنارة للاستشارات
www.manaraa.com

6 summarizes my research and offers broad implications and future directions for both chemical education practitioners and researchers.

1.4 References

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

Bhattacharyya, G., (2013), From source to sink: mechanistic reasoning using the electron-pushing formalism. *J. Chem. Educ.*, **90**, 1282–1289.

Bhattacharyya, G., Bodner, G.M., (2005), "It gets me to the product": how students propose organic mechanisms. *J. Chem. Educ.*, **82**, 1402–1407.

Bodé, N.E., Deng, J.M., Flynn, A.B., (2019), Getting past the rules and to the WHY: Causal mechanistic arguments when judging the plausibility of organic reaction mechanisms. *J. Chem. Educ.*, **96**, 1068–1082.

Caspari, I., Kranz, D., Graulich, N., (2018a), Resolving the complexity of organic chemistry students' reasoning through the lens of a mechanistic framework. *Chem. Educ. Res. Pract.*, **19**, 1117–1141.

Caspari, I., Weinrich, M.L., Sevian, H., Graulich, N., (2018b), This mechanistic step is "productive": organic chemistry students' backward-oriented reasoning. *Chem. Educ. Res. Pract.*, **19**, 42–59.

Cooper, M., Klymkowsky, M., (2013), Chemistry, Life, the Universe, and Everything: A new approach to general chemistry, and a model for curriculum reform. *J. Chem. Educ.*, **90**, 1116–1122.

Cooper, M., Kouyoumdjian, H., Underwood, S., (2016), Investigating students' reasoning about acid–base reactions. *J. Chem. Educ.*, **93**, 1703–1712.

Cooper, M., Stowe, R.L., Crandell, O.M., Klymkowsky, M.W., (2019), Organic Chemistry, Life, the Universe and Everything (OCLUE): A transformed organic chemistry curriculum. *J. Chem. Educ.*, **96**, 1858–1872.

Crandell, O., Kouyoumdjian, H., Underwood, S., Cooper, M., (2018), Reasoning about reactions in organic chemistry: starting It in general chemistry. *J. Chem. Educ.*, **96**, 213–226.

Cruz-Ramírez de Arellano, D., Towns, M.H., (2014), Students' understanding of alkyl halide reactions in undergraduate organic chemistry. *Chem. Educ. Res. Pract.*, **15**, 501–515.

Flynn, A.B., Featherstone, R.B., (2017), Language of mechanisms: exam analysis reveals students' strengths, strategies, and errors when using the electron-pushing formalism (curved arrows) in new reactions. *Chem. Educ. Res. Pract.*, **18**, 64–77.

Flynn, A.B., Ogilvie, W.W., (2015), Mechanisms before reactions: A mechanistic approach to the organic chemistry curriculum based on patterns of electron flow. *J. Chem. Educ.*, **92**, 803–810.

Galloway, K.R., Stoyanovich, C., Flynn, A.B., (2017), Students' interpretations of mechanistic language in organic chemistry before learning reactions. *Chem Educ Res Pract.*, **18**, 353–374.

Grove, N., Cooper, M., Rush, K., (2012), Decorating with arrows: Toward the development of representational competence in organic chemistry. *J. Chem. Educ.*, **89**, 844–849.

Ha, M., Nehm, R.H., Urban-Lurain, M., Merrill, J.E., (2011), Applying computerized-scoring models of written biological explanations across courses and colleges: prospects and limitations. *CBE-Life Sci. Educ.*, **10**, 379–393.

المنارة للاستشارات
www.manaraa.com

Haudek, K.C., Kaplan, J.J., Knight, J., Long, T., Merrill, J., Munn, A., Nehm, R., Smith, M., Urban-Lurain, M., (2011), Harnessing technology to improve formative assessment of student conceptions in STEM: forging a national network. *CBE-Life Sci. Educ.*, **10**, 149–155.

Haudek, K.C., Prevost, L.B., Moscarella, R.A., Merrill, J., Urban-Lurain, M., (2012), What are they thinking? Automated analysis of student writing about acid−base chemistry in introductory biology. *CBE Life Sci. Educ.*, **11**, 283–293.

Kaplan, J.J., Haudek, K.C., Ha, M., Rogness, N., Fisher, D.G., (2014), Using lexical analysis software to assess student writing in statistics. *Technol. Innov. Stat. Educ.*, **8**.

Kermack, W.O., Robinson, R., (1922), LI.—An explanation of the property of induced polarity of atoms and an interpretation of the theory of partial valencies on an electronic basis. *J. Chem. Soc. Trans.*, **121**, 427–440.

Kozma, R., (2003), The material features of multiple representations and their cognitive and social affordances for science understanding. *Learn. Instr.,* External and Internal Representations in Multimedia Learning **13**, 205–226.

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

Liu, X., (2015), Applied Ordinal Logistic Regression Using Stata: From Single-Level to Multilevel Modeling. SAGE Publications.

McNemar, Q., (1947), Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, **12**, 153–157.

Morrison, R.T., Boyd, R.N., (1959), Organic Chemistry. Allyn & Bacon, Boston, MA.

National Research Council, (2011), A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas.

Popping, R., (2000), Computer-Assisted Text Analysis. SAGE.

Prevost, L.B., Haudek, K., Urban-Lurain, M., Merrill, J., (2012), Examining student constructed explanations of thermodynamics using lexical analysis, in: 2012 Frontiers in Education Conference Proceedings. Presented at the 2012 Frontiers in Education Conference Proceedings, pp. 1–6.

Prevost, L.B., Haudek, K.C., Norton Henry, E., Urban-Lurain, M., Berry, M.C., (2013), Automated text analysis facilitates using written formative assessments for just-in-time teaching in large enrollment courses. Presented at the 120th ASEE Annual Conference & Exposition.

Prevost, L.B., Smith, M.K., Knight, J.K., (2016), Using student writing and lexical analysis to reveal student thinking about the role of stop codons in the central dogma. *CBE-Life Sci. Educ.*, **15**, ar65: 1-13.

Sheskin, D.J., (2011), Test 20: The McNemar Test, in: Handbook of Parametric and Nonparametric Statistical Procedures. Chapman & Hall/CRC, Boca Raton, FL, pp. 835–882.

Strickland, A.M., Kraft, A., Bhattacharyya, G., (2010), What happens when representations fail to represent? Graduate students' mental models of organic chemistry diagrams. *Chem. Educ. Res. Pract.*, **11**, 293–301.

Webber, D.M., Flynn, A.B., (2018), How are students solving familiar and unfamiliar organic chemistry mechanism questions in a new curriculum? *J. Chem. Educ.*, **9**, 1451-1467.

Weinrich, M.L., Sevian, H., (2017), Capturing students' abstraction while solving organic reaction mechanism problems across a semester. *Chem. Educ. Res. Pract.*, **18**, 169–190.

Weston, M., Haudek, K.C., Prevost, L., Urban-Lurain, M., Merrill, J., (2015), Examining the impact of question surface features on students' answers to constructed-response questions on photosynthesis. *CBE-Life Sci. Educ.*, **14**, ar19: 1-12.

Wheeler, D.M.S., Wheeler, M.M., (1982), Trends in the teaching of organic chemistry: A survey of some textbooks. *J. Chem. Educ.*, **59**, 863-865.

المنارة للاستشارات
www.manaraa.com

Chapter 2

Methods

This chapter will detail the methods used for the work presented in this dissertation. It will cover interrater reliability, logistic regression, and the McNemar test. Much of this work was completed with the assistance of a Python program, which will also be described in depth in this chapter.

2.1 Interrater reliability (IRR)

All of the work presented in this dissertation relies heavily on qualitative coding of students' written responses. In order to add trustworthiness to a study, multiple raters code the same set of data and compare their codes to check agreement. When multiple raters code the same data set, interrater reliability (IRR) can be established (Lange, 2011). These raters can be two different humans, or in the case of much of the work presented in this dissertation, one human and one predictive model. There are many different options for measuring the association between two different raters, including percent agreement and Cohen's kappa (Cohen, 1960).

2.1.1 Percent agreement

One measure of interrater reliability is percent agreement. Percent agreement is the raw percentage of the number of observations coders agree on and can range from 0 to 100 (Lombard *et al.*, 2002). This can accommodate any number of coders. Percent agreement is simple and quick to calculate and is easily

المنارة للاستشارات
www.manaraa.com

interpretable. Percent agreement is used as the measure of IRR between human coding and computer coding throughout this dissertation.

2.1.2 Cohen's kappa

Another method of measuring IRR is using Cohen's kappa. Kappa is used to correct for any agreement between two raters that may have occurred by chance (Cohen, 1960). The measure can be used when the data is categorical, the categories are mutually exclusive, the observations are paired (i.e., you are comparing the rating of two raters for the same observation), the cross-tabulation is symmetric, the raters are independent, and the same two raters evaluate all observations (Cohen, 1960). Kappa values range from -1 to +1; values below 0 indicate that the agreement between raters is lower than would be expected by chance (Cohen, 1960). According to Cohen, the value of kappa should be interpreted as follows: 0.01-0.20 as none to slight agreement, 0.21-0.40 as fair agreement, 0.41-0.60 as moderate agreement, 0.61-0.80 as substantial agreement, and 0.8-1.00 as almost perfect agreement. However, it has been suggested that interpretation of kappa should be less lenient, with no agreement at 0-0.20, minimal agreement at 0.21-0.39, weak agreement at 0.40-0.59, moderate agreement at 0.60-0.79, strong agreement at 0.80-0.90, and almost perfect agreement above 0.90 (McHugh, 2012).

2.1.3 IRR in this work

The published studies in Chapters 3 and 5 use percent agreement as the measure of accuracy for the predictive models. Cohen's kappa was not used. However, I report kappa here. Chapter 3 contains a training set and a testing set. The training set was 87% accurate with a kappa value of 0.78. The testing set was 82% accurate with a kappa value of 0.69. According to Cohen's interpretation recommendations for kappa, both sets have substantial agreement and according to the McHugh's more stringent interpretation recommendations, both data sets have moderate agreement. Chapter 5 also contains training and testing sets.

المنارة للاستشارات
www.manaraa.com

The training set was 92% accurate with a kappa value of 0.86. The testing set was 87% accurate with a kappa value of 0.77. According to Cohen's recommendations, the training set has almost perfect agreement and the testing set has substantial agreement (Cohen, 1960). McHugh's recommendations classify the training set as strong agreement and the testing set as moderate agreement (McHugh, 2012). Thus, the inclusion of Cohen's kappa is consistent with the conclusions of the work: the predictive models are an acceptable and accurate way to score student responses for formative assessments.

2.2 Logistic regression

Logistic regression is a procedure used to predict group membership using predictor variables (Cox, 1958; DeMaris, 2003; Hosmer Jr., 2013; Sheskin, 2011a). Logistic regression involves one dependent variable and any number of independent variables (Cox, 1958). The procedure has been used previously in chemistry education research (e.g., Lewis, 2014; Srinivasan *et al.*, 2018; Tang *et al.*, 2014) and for predicting an outcome to score student text responses (e.g., Prevost *et al.*, 2016). The null hypothesis of the test states that there is no relationship between the dependent variable and the predictor variables (i.e., the predictor variables do not predict the category of the dependent variable; Cox, 1958). The predictor variables can be continuous, categorical, or a combination of the two (Sheskin, 2011a). Logistic regression can be binomial (i.e., there are only two possible outcomes for the dependent variable; Cox, 1958) or multinomial (i.e., there are more than two possible outcomes for the dependent variable; Engel, 1988). Logistic regression uses a logit function which is the natural logarithm of the odds of an observation belonging to a particular outcome category (Cox, 1958; Sheskin, 2011a):

$$logitp(x) = ln(\frac{p(x)}{1-p(x)}) = \beta_0 + \beta_n x \tag{2.1}$$

المنارة للاستشارات
www.manaraa.com

p(x) represents the probability that the dependent variable will be a certain outcome (e.g., using the Lewis acid−base model) while β_0 is the constant and β_n is the regression coefficient for *n* predictor variable (Sheskin, 2011). The regression coefficient is "the estimated increase in the log odds of the outcome per unit increase in the value of the [predictor variable]" (Szumilas, 2010).

An odds ratio is defined as the probability of an outcome occurring over the probability of an outcome not occurring (Szumilas, 2010). It is a measure of the odds that a certain outcome will occur if a certain predictor variable increases by one unit (i.e., is present for categorical variables) compared to the odds of the outcome occurring if the predictor variable decreases by one unit (i.e., is not present for categorical variables) when holding the values of the other predictor variables constant (Ranganathan *et al.*, 2017). The odds ratio for a predictor variable is calculated by taking the exponential function of the regression coefficient. An odds ratio that is greater than one indicates that the odds of the outcome occurring are higher when the predictor variable is included and an odds ratio between zero and one indicates that the odds of the outcome occurring are lower when the predictor variable is included. One recommendation for interpreting the magnitude of odds ratios comes from Chen *et al.* (2010) who found that in epidemiological studies, odds ratios of 1.68, 3.47, and 6.71 relate to Cohen's *d* (Cohen, 1988) effect sizes of 0.2 (small), 0.5 (medium), and 0.8 (large), respectively.

One advantage to using logistic regression over linear regression is that logistic regression need not follow assumptions of normality for predictor variables and only requires that the observations in an analysis are independent of one another (Tabachnick and Fidell, 2001). Additionally, logistic regression assumes the dependent and independent variable have a nonlinear relationship (i.e., a logarithmic relationship; Sheskin, 2011a). This is ideal over linear regression because logistic regression will only predict probabilities between 0 and 1, while linear regression can predict probabilities less than 0 and greater than 1, which are not statistically possible (Feller, 1950).

المنارة للاستشارات
www.manaraa.com

Chapters 3 and 5 use binomial logistic regression models to predict outcome variables by using a split-sample as means of validation of the model (Picard and Berk, 1990; Picard and Cook, 1984; Snee, 1977). This is done by separating the dataset into a training set and a testing set, sometimes called a modeling set and an external evaluation set, respectively (e.g., Martin *et al.*, 2012). The testing set is used to develop a predictive logistic regression model and the testing set is used to test the model once it has been developed (Snee, 1977). The ratio of training to testing set splits varies, typically from 50:50 (Snee, 1977) to 90:10 (Steyerberg *et al.,* 2001). For the data in this dissertation, the data are split 70:30 (e.g., Prevost *et al.*, 2016) to ensure the training set is of ample size for model development while still leaving enough data to test the model on a dataset approximately the size of a large lecture organic chemistry course. Each logistic regression model was developed through an iterative process of making small changes to the predictor variable categories by looking at the incorrect predictions in the training set in order to increase predictive accuracy. It is up to the researcher to decide when to stop when they are no longer able to make the model significantly more accurate by changing the categories. Once the model is finalized, it can be run on the testing set and prediction accuracy can be calculated (e.g., Prevost *et al.*, 2016).

In Chapter 3, a binomial logistic regression model is used to predict use or non-use of the Lewis acid−base model when describing and explaining an acid−base proton transfer reaction mechanism. Use is coded as *1* and non-use is coded as *0*. Multinomial logistic regression techniques can be used to predict a dependent variable that has three or more categories and ordinal logistic regression can be used when multiple categories are ordered, but using two binomial logistic regression models was more accurate in the case of the study in Chapter 5, so multinomial and ordinal logistic regression were not used. In Chapter 5, two binomial logistic regression models are used and combined to result in a tri-level score prediction. Model #1 predicts membership in Levels 2 or 3 (*1*) as opposed to Level 1 (*0*). Model #2 predicts membership in Level 3 (*1*) as opposed to Levels 1 and 2 (*0*).

المنارة للاستشارات
www.manaraa.com

In this study, the purpose of using logistic regression analysis was to generate as accurate of a predictive model as possible by iteratively editing the predictor variables (i.e., categories; Stahel, 2004). Thus, it is inappropriate to overanalyze the coefficients and odds ratios of the predictors because the predictors were specifically edited to achieve the highest accuracy possible with the model. In Chapters 3 and 5, there is some commentary on the significance and magnitude of predictor variables. This was included because reviewers were interested in seeing such commentary. Though the significance and magnitude of predictor variables can be informative, one should be careful not to overinterpret these magnitudes in this case because the main goal of the logistic regression model was to generate accurate predictions.

2.3 McNemar test

The McNemar test is a statistical test used to see if two dependent samples represent two independent populations (McNemar, 1947; Sheskin, 2011b). It can be used to test the impact of a treatment on a group of people by testing for a dichotomous variable before and after treatment without the need for a control group. The test is preferred in clinical settings when having a treatment group may be impractical or unethical (Liao and Lin, 2008). The McNemar test is used to monitor behavior and attitudes (e.g., Fava *et al.*, 2002; Flisher *et al.*, 2004; MacKenzie *et al.*, 1996) and in evaluating the effectiveness of medical treatments (Ahn *et al.*, 1999; Caronni and Sciumé, 2017; Keid *et al.*, 2009; Sarosdy *et al.*, 1995; van der Poel *et al.*, 1998). The test assumes that there is one categorical variable with two categories and one independent variable with two paired samples, the two categories in the dependent variable are mutually exclusive, and the sample is random (Sheskin, 2011b). The null hypothesis states that the means of the paired samples are equal (McNemar, 1947).

For example, in Chapter 4 I evaluate the impact of a tutorial designed to increase student usage of the Lewis acid−base model when responding to an item asking what is happening in an acid−base proton

المنارة للاستشارات
www.manaraa.com

transfer reaction mechanism and why. Rather than only providing only some students with a tutorial in order to have a control or using a more complex experimental design to provide all students the tutorial, I opted to provide everyone with the same treatment and employ the McNemar test for evaluation. Responses were evaluated for use or non-use of the Lewis model before and after the tutorial. This produced four groups of students: students who used the Lewis model before and after the tutorial (a), students who used the Lewis model before the tutorial and did not use the Lewis model after the tutorial (b), students who did not use the Lewis model before the tutorial but did use the Lewis model after the tutorial (c), and students who did not use the Lewis model before or after the tutorial (d). These four groups are shown in tabular form in Table 2.1.

Table 2.1 The four groups of a McNemar test.

	Post-test	
Pre-test	Lewis "use"	Lewis "non-use"
Lewis "use"	a	b
Lewis "non-use"	c	d

The ideal outcome is for students to use the Lewis model after the tutorial and the non-ideal outcome is for students to not use the Lewis model after the tutorial. The test statistic is a χ^2 value which is computed using the number of cases which move from the non-ideal outcome to the ideal outcome (i.e., b) after the tutorial and the number of cases which move from the ideal outcome to the non-ideal outcome after the tutorial (i.e., c). The test statistic is computed using the following equation:

$$\chi^2 = \frac{(b-c)^2}{b+c} \tag{2.2}$$

Depending on the value of χ^2, the null hypothesis can be rejected or retained. Like with logistic regression, odds ratios can be calculated for a McNemar test for further interpretation. An odds ratio of

المنارة للاستشارات
www.manaraa.com

greater than one indicates a positive impact of the treatment on achieving the ideal outcome and an odds ratio between zero and one indicates a negative impact of the treatment on achieving the ideal outcome. The odds ratio is calculated by dividing the number of cases which move from the non-ideal outcome to the ideal outcome after the tutorial (i.e., b) by the number of cases which move from the ideal outcome to the non-ideal outcome after the tutorial (i.e., c; Cleophas and Zwinderman, 2016) . For example, if 103 students moved from not using the Lewis model before the tutorial to using the Lewis model after the tutorial and only 13 students moved from using the Lewis model before the tutorial to not using the Lewis model after the tutorial, the odds ratio would be 103/13 or 7.92. This can be interpreted as follows: the odds of a person moving from non-use of the Lewis acid−base model to use of the Lewis acid−base model after the tutorial are 7.92 times higher than the odds of a person moving from use of the Lewis acid−base model before the tutorial to non-use of the Lewis acid−base model after the tutorial (Cleophas and Zwinderman, 2016; McNemar, 1947). This would be a large effect size according to the suggestions for interpreting odds ratios reported above (Chen *et al.*, 2010).

2.4 Python program

SPSS Modeler (IBM Corp., 2017) was originally used for lexical analysis and development of predictive logistic regression models. However, several limitations exist with the program that were undesirable for meeting the goals of this project. For example, the software is expensive and creates a barrier for making the predictive models freely available. A Python program (van Rossum and Drake, 2011) was developed to assign categories to written responses and create predictive logistic regression models.

The program first reads in a .csv file containing a list of responses and their scores as assigned by a human coder. Then there are four programs that prepare and analyze the responses for the predictive model. These programs run in succession: equivalents, types, rules, and categories. Each program reads in a .csv

المنارة للاستشارات
www.manaraa.com

file with information that has been added by the researcher about what to do with the responses that have been input.

The first program is called Equivalents. Equivalents are words or phrases that are exactly the same in the context of the responses. These can include misspellings and different ways to write something. For example, the equivalent “alcohol” includes “OH” because, in the context of the reaction mechanisms provided, those two words mean the same thing. This equivalent could also include misspellings or typos like “aclohol”. The equivalents program essentially cleans the data and replaces all equivalents with the same term (in this case, “alcohol”). Once this program has run, the data that is output by the program can be run through the next program, Types.

Types are groups of words that are synonyms. The words are grouped together and replaced by the name of the type. The program reads in a .csv file which includes information about what words and phrases belong in each type. One example of a type would be *halogen*. This type would include the terms “bromine”, “chlorine”, and “iodine”. In the case of the reaction mechanism items in this dissertation, halogens are varied as leaving groups for different iterations of the items. Thus, the words are synonyms in the context of the data set. All words that are included in the type are replaced by their type in the text. Once these replacements occur, the Rules program can be run on the data.

Rules are text patterns that are made up of two or more words, phrases, equivalents, and types in a particular order, within six or less words of each other. The program reads in information about rules from a .csv file. An example of a rule is *attract positive.* To determine if a response includes this rule, the program looks for the types *attract* and *positive* in that order and within six words of each other. An example phrase that would fit this rule is “the alcohol is **attracted** to the **positively** charged carbocation.” The word “attracted” is included in the type *attract* and the word positively is included in the type *positive*, so the rule finds the two types within six words of each other. Multiple rules can be developed if both word orders are desired (i.e., *attract positive* and *positive attract*). An example of the same sentiment but different

المنارة للاستشارات
www.manaraa.com

word order is “the **positively** charged carbocation **attracts** the alcohol.” These two rules could be added to the same category by the category program. The ability to change the word order for rules has the potential to distinguish between cases of understanding and misunderstanding for students based on text patterns, but the models in this study did not score for correctness. Rules do not make replacements in the text, meaning that they are not mutually exclusive and one instance of a type can fit multiple rules.

Finally, the Categories program is run. Categories are combinations of specific words, phrases, rules, or types found in responses. Categories are the predictor variables that go into the logistic regression model. An example of a category is *bond forming*. The category includes the phrases “new bond” and two rules about bond formation: *form bond* and *bond form*. Categories can also look for the absence of certain words, phrases, rules, or types. An example of this is the category *absence of explanation* which includes responses that omit many words and types that indicate a student is addressing the why behind the reaction (e.g., the types *attraction* and *electronegativity*). The program reads in the categories that have been input by the researcher in a .csv file and searches for the words, phrases, rules, and types that are in each category. If a response is positive for any of the words, phrases, rules, or types for a specific category, the program gives that response a *1* for the category. One response can contain any number of categories, from no categories to all of the categories. Whether or not a response is positive for the different categories is what determines the prediction by the logistic regression model.

In order to use the response categories to develop and apply a predictive logistic regression model, code was adapted from Pedregosa *et al.* (2011). The code is used to develop the model using the categories of the training set and then used to apply the model to the testing set. The program outputs level predictions, which can then be compared to the human codes for each response to determine accuracy. In the study in Chapter 5, two binomial logistic regression models are used and combined to determine the overall level of each response (i.e., Level 1, 2, or 3).

المنارة للاستشارات
www.manaraa.com

In order to quickly determine overall levels from the two combined binomial logistic regression models, an additional program was written to combine the models and determine the overall level for each response. The program takes the output from the predictions made by Model #1 (Level 1 is *0* and Level 2/3 is *1*) and Model #2 (Level 1/2 is *0* and Level 3 is *1*) and recodes all responses to their overall level. All responses predicted to be *0* by Model #1 are predicted to be Level 1 and all responses coded as *1* by Model #2 are predicted to be Level 3. Responses that are coded as *1* by Model #1 and *0* by Model #2 are predicted to be Level 2.

2.5 References

Ahn, C.W., Song, Y.D., Kim, J.H., Lim, S.K., Choi, K.H., Kim, K.R., Lee, H.C., Huh, K.B., (1999), The validity of random urine specimen albumin measurement as a screening test for diabetic nephropathy. *Yonsei Med. J.*, **40**, 40–45.

Caronni, A., Sciumè, L., (2017), Is my patient actually getting better? Application of the McNemar test for demonstrating the change at a single subject level. *Disabil. Rehabil.*, **39**, 1341–1347.

Chen, H., Cohen, P., Chen, S., (2010), How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Commun. Stat. - Simul. Comput.*, **39**, 860–864.

Cleophas, T.J., Zwinderman, A.H., (2016), McNemar's Odds Ratios, in: Cleophas, T.J., Zwinderman, A.H. (Eds.), Clinical Data Analysis on a Pocket Calculator: Understanding the Scientific Methods of Statistical Reasoning and Hypothesis Testing. Springer International Publishing, Cham, pp. 275–278.

Cohen, J., (1988), Statistical power analysis for the behavioral sciences. L. Erlbaum Associates, Hillsdale, N.J.

Cohen, J., (1960), A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.*, **20**, 37–46.

Cox, D.R., (1958), The regression analysis of binary sequences. *J. R. Stat. Soc. Ser. B Methodol.*, **20**, 215–242.

DeMaris, A., (2003), Logistic Regression, in: Handbook of Psychology. American Cancer Society, pp. 509–532.

Engel, J., (1988), Polytomous logistic regression. *Stat. Neerlandica*, **42**, 233–252.

Fava, M., Farabaugh, A.H., Sickinger, A.H., Wright, E., Alpert, J.E., Sonawalla, S., Nierenberg, A.A., Iii, J.J.W., (2002), Personality disorders and depression. *Psychol. Med.*, **32**, 1049–1057.

Feller, W., (1950), An Introduction to Probability Theory and Its Applications. Wiley.

Flisher, A.J., Evans, J., Muller, M., Lombard, C., (2004), Brief report: Test–retest reliability of self-reported adolescent risk behaviour. *J. Adolesc.*, **27**, 207–212.

Hosmer Jr., D.W., Lemeshow, S., Sturdivant, R.X., (2013), Applied Logistic Regression. John Wiley & Sons.

IBM Corp. Released June 2017. IBM SPSS Modeler for Mac OS X, Version 18.1. Armonk, NY.

المنارة للاستشارات
www.manaraa.com

Keid, L.B., Soares, R.M., Vasconcellos, S.A., Megid, J., Salgado, V.R., Richtzenhain, L.J., (2009), Comparison of agar gel immunodiffusion test, rapid slide agglutination test, microbiological culture and PCR for the diagnosis of canine brucellosis. *Res. Vet. Sci.*, **86**, 22–26.

Lange, R.T., (2011), Inter-rater Reliability, in: Kreutzer, J.S., DeLuca, J., Caplan, B. (Eds.), Encyclopedia of Clinical Neuropsychology. Springer, New York, NY, pp. 1348–1348.

Lewis, S.E., (2014), An Introduction to Nonparametric Statistics in Chemistry Education Research, in: Tools of Chemistry Education Research, ACS Symposium Series. American Chemical Society, pp. 115–133.

Liao, Y.-Y., Lin, Y.-M., (2008), McNemar Test Is Preferred for Comparison of Diagnostic Techniques. *Am. J. Roentgenol.*, **191**, W190–W190.

Lombard, M., Snyder-Duch, J., Bracken, C.C., (2002), Content analysis in mass communication: assessment and reporting of intercoder reliability. *Hum. Commun. Res.*, **28**, 587–604.

MacKenzie, D.M., Copp, P., Shaw, R.J., Goodwin, G.M., (1996), Brief cognitive screening of the elderly: a comparison of the Mini-Mental State Examination (MMSE), Abbreviated Mental Test (AMT) and Mental Status Questionnaire (MSQ). *Psychol. Med.*, **26**, 427–430.

Martin, T.M., Harten, P., Young, D.M., Muratov, E.N., Golbraikh, A., Zhu, H., Tropsha, A., (2012), Does rational selection of training and test sets improve the outcome of QSAR modeling? *J. Chem. Inf. Model.*, **52**, 2570–2578.

McHugh, M.L., (2012), Interrater reliability: the kappa statistic. *Biochem. Medica*, **22**, 276–282.

McNemar, Q., (1947), Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, **12**, 153–157.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., (2011), Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, **12**, 2825-2830.

Picard, R.R., Berk, K.N., (1990), Data splitting. *Am. Stat.*, **44**, 140–147.

Picard, R.R., Cook, R.D., (1984), Cross-validation of regression models. *J. Am. Stat. Assoc.*, **79**, 575–583.

Prevost, L.B., Smith, M.K., Knight, J.K., (2016), Using student writing and lexical analysis to reveal student thinking about the role of stop codons in the central dogma. *CBE-Life Sci. Educ.*, **15**, ar65: 1-13.

Ranganathan, P., Pramesh, C.S., Aggarwal, R., (2017), Common pitfalls in statistical analysis: Logistic regression. *Perspect. Clin. Res.*, **8**, 148–151.

Sarosdy M. F., deVere White R.W., Soloway M.S., Sheinfeld J., Hudson M.A., Schellhammer P.F., Jarowenko M.V., Adams G., Blumenstein B.A., (1995), Results of a multicenter trial using the BTA test to monitor for and diagnose recurrent bladder cancer. *J. Urol.*, **154**, 379–384.

Sheskin, D., (2011a), Test 39: Logistic Regression, in: Handbook of Parametric and Nonparametric Statistical Procedures. Chapman & Hall/CRC, pp. 1593–1625.

Sheskin, D., (2011b), Test 20: The McNemar Test, in: Handbook of Parametric and Nonparametric Statistical Procedures. Chapman & Hall/CRC, Boca Raton, FL, pp. 835–882.

Snee, R.D., (1977), Validation of regression models: Methods and examples. *Technometrics*, **19**, 415–428.

Srinivasan, S., Gibbons, R.E., Murphy, K.L., Raker, J., (2018), Flipped classroom use in chemistry education: results from a survey of postsecondary faculty members. *Chem. Educ. Res. Pract.*, **19**, 1307–1318.

Stahel, W.A., (2004), Purposes and strategies in regression analysis. *J. Stat. Plan.* Inference, Contemporary Data Analysis: Theory and Methods in **122**, 175–186.

Steyerberg, E.W., Harrell, F.E., Borsboom, G.J.J.M., Eijkemans, M.J.C., Vergouwe, Y., Habbema, J.D.F., (2001), Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *J. Clin. Epidemiol.*, **54**, 774–781.

Szumilas, M., (2010), Explaining odds ratios. *J. Can. Acad. Child Adolesc. Psychiatry*, **19**, 227–229.

المنارة للاستشارات
www.manaraa.com

Tabachnick, B.G., Fidell, L.S., (2001), Using Multivariate Statistics. Allyn and Bacon.

Tang, H., Kirk, J., Pienta, N.J., (2014), Investigating the effect of complexity factors in stoichiometry problems using logistic regression and eye tracking. *J. Chem. Educ.*, **91**, 969–975.

van der Poel, H.G., van Balken, M.R., Schamhart, D.H.J., Peelen, P., de Reijke, T., Debruyne, F.M.J., Schalken, J.A., Witjes, J.A., (1998), Bladder wash cytology, quantitative cytology, and the qualitative BTA test in patients with superficial bladder cancer. *Urology*, **51**, 44–50.

van Rossum, G., Drake, F.L., (2011), The Python Language Reference Manual. Network Theory Ltd.

www.manaraa.com

Chapter 3

Using lexical analysis to predict Lewis acid−base model use in responses to an acid−base proton-transfer reaction

3.1 Note to Reader

This chapter is a published manuscript in the *Journal of Chemical Education*, reprinted with permission from:

Dood, A. J., Fields, K. B., & Raker, J. R. (2018) Using lexical analysis to predict Lewis acid−base model use in responses to an acid−base proton-transfer reaction. *Journal of Chemical Education*, 95(8) 1267-1275. DOI: 10.1021/acs.jchemed.8b00177.

Copyright 2018 American Chemical Society.

Further information regarding permissions can be found in Appendix C.

This work was published with co-authors. Kimberly B. Fields is an organic chemistry instructor at the University of South Florida who allowed for data collection in her courses. Jeffrey R. Raker is the principal investigator for this project.

3.2 Abstract

The Lewis acid−base model is key to identifying and explaining the formation and breaking of bonds in a large number of reaction mechanisms taught in the sophomore-level yearlong organic chemistry course.

المنارة للاستشارات
www.manaraa.com

Understanding the model is, thus, essential to success in organic chemistry coursework. Concept inventories exist to identify misunderstandings and misconceptions of acid–base theories; open-ended problems, though, have been shown to provide a more nuanced and holistic understanding of how students use acid–base models to explain reactions. The time necessary to score such problems, however, limits their use, especially in large student enrollment courses. Given the efficacy of open-ended problems, there is occasion for the development of methods to efficiently and effectively analyze open-ended assessment responses. In this study, we establish the importance of assessing "use of the Lewis acid–base model to explain a chemical reaction" by determining the association of model use with summative examination performance. In addition, we generate and evaluate a binomial logistic regression model based on lexical analysis techniques for predicting Lewis acid–base model use in explanations of an acid–base proton-transfer reaction. Our work results in a predictive model that can be used to score the open-ended problem used in our study.

3.3 Graphical abstract

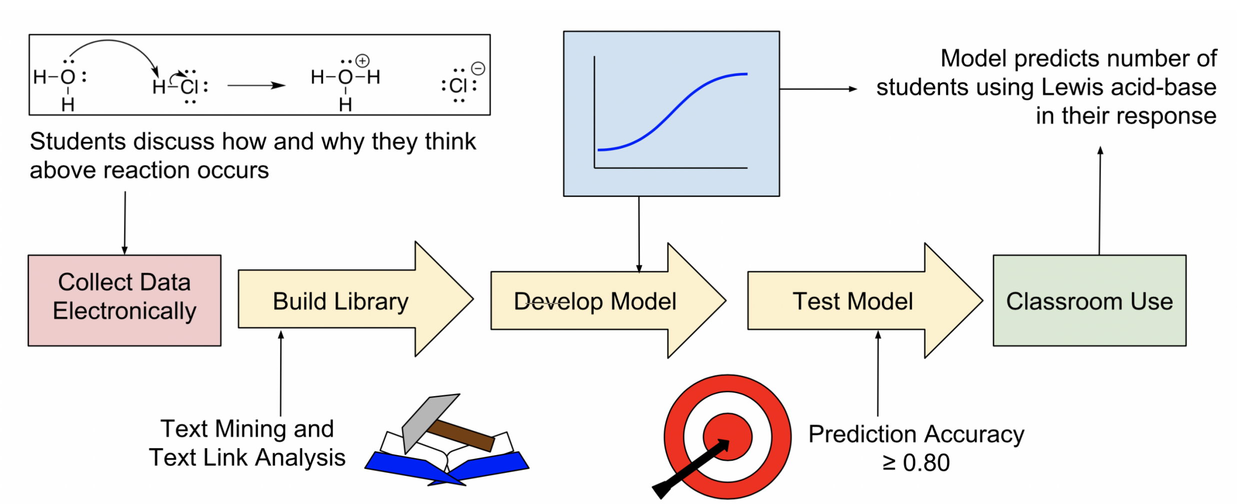


Figure 3.1. Graphical abstract for Chapter 3

www.manaraa.com

3.4 Introduction

Formative assessment is critical to instructors gaining valuable insight into student understanding during the learning experience (Black and Wiliam, 1998; National Research Council, 2001). Use of pre-lecture quizzes (Kristine, 1985; Seery, 2012; Seery and Donnelly, 2012) or during-lecture classroom response systems (Gebru *et al.*, 2012; King, 2011; MacArthur and Jones, 2008) are key examples of formative assessment. Formative assessment tools rely on multi-choice, numeric, and small, often single word answers that can be quickly and easily scored, especially for large numbers of students (Caldwell, 2007). However, not all tools are conducive to measuring all types of knowledge and skills, nor are they all conducive to easy customization. For example, the development of distractor responses for concept-inventories is resource-intensive (Bretz and Linenberger, 2012; Christian and Yezierski, 2012); changing an item prompt from explaining the dissociation of a monoprotic acid to the dissociation of a diprotic acid has implications for what rationales a respondent might give in an open-ended response environment, thus there is a need to revise the distractor set before use. While efforts could be devoted to building an array of multiple-choice assessment items that evaluate acid−base understanding, the development of tools that effectively and efficiently analyze a wide-range of open-ended responses to assessments may be more appropriate for assessing student understanding. We report herein the modification of a formative assessment item shown to evaluate model use in explaining an acid−base reaction (Cooper *et al.*, 2016). We find that students who used a Lewis acid−base model on the assessment item score higher on acid−base summative assessment items. Lastly, we report a predictive logistic regression model based on lexical analysis techniques that can be used to evaluate open-ended responses to the modified assessment item; the predictive logistic regression model can then be used by instructors to quickly probe student understanding in a formative assessment context.

www.manaraa.com

3.4.1 Student understanding of acid−base models

Students struggle to understand and apply acid−base models, especially the Lewis acid−base model (Cartrette and Mayo, 2011). Many authors have stated that advanced understanding of the Lewis acid−base model is critical to success in organic chemistry and upper level courses; however, this claim has not been well substantiated with empirical data (Bhattacharyya, 2013; Cooper *et al.*, 2016; Nataro *et al.*, 2004; Stoyanovich *et al.*, 2015). The struggle to understand and apply acid−base models has been observed in a multitude of courses across the undergraduate and graduate chemistry curriculum (Bhattacharyya, 2006; Bretz and McClary, 2015; McClary and Talanquer, 2011a; Stoyanovich *et al.*, 2015). A small-scale survey found that organic chemistry instructors view acid−base chemistry as a useful topic in their courses and for courses later in the undergraduate chemistry curriculum (Duis, 2011). There are three main models: Arrhenius, Brønsted-Lowry, and Lewis (Paik, 2015). Confusion with acid−base concepts has been hypothesized to be due to unclear explanations of the relationship between these three models (Friesen, 2008; Paik, 2015). While emphasis is placed on the Brønsted-Lowry model in general chemistry, the Lewis model has broader applicability across the curriculum (Herron, 1953; Luder, 1948; Shaffer, 2006). Cartrette and Mayo (2011) found that undergraduate students could correctly use the Brønsted-Lowry model on a series of tasks, but less than half of the students in their study could correctly use a Lewis model. Cooper *et al.* (2016) found that students who used the Lewis model to explain an acid−base proton-transfer reaction were more likely to correctly produce its mechanism. While a well-developed understanding of the Lewis acid−base model is important for learning organic chemistry, use of the Lewis model in explaining acid−base reactions has not been directly evaluated against performance in the course.

3.4.2 Tools to measure understanding of acid−base models

Assessment of acid−base understanding and associated research studies heavily rely on in-depth interviews and multiple-choice items (or sets of multiple-choice items, e.g., concept-inventories). On one end

المنارة للاستشارات
www.manaraa.com

of the spectrum are open-ended, semi-structured interviews that require lengthy development, administration, and analysis; such interviews, indicative of oral assessments (Dicks *et al.*, 2012; Roecker, 2007), provide the most comprehensive depth of evaluation of a student's understanding of acids and bases. McClary and Talanquer (2011a) reported the results of an interview-based study into student conceptions of acids and bases; they found that their participants used several models to describe acid and base strength, and that some participants used mixed models or changed their mental models based on the nature of the tasks given. Interviews and oral assessments, however, are ultimately not pragmatic for use by an instructor in the context of a course except in the case of small enrollment courses.

At the other end of the spectrum are multiple-choice based assessment items and concept-inventories (i.e., a group of multiple-choice items). An accepted method for developing concept-inventories includes a broad interview study or large collection of open-ended responses from which appropriate distractors are developed and evaluated. The process includes evaluating the ability of the assessment item or items to produce valid and reliable measures of understanding (Arjoon *et al.*, 2013). Studies rely on these tools to assess student understanding and alternative conceptions of concepts in organic chemistry (Bretz and McClary, 2015; Henderleiter *et al.*, 2001; Rushton *et al.*, 2008), with instruments developed to assess students' critical thinking pathways (Taagepera and Noori, 2000), understanding of representations (Domin and Bodner, 2012), hierarchical levels of learning (Hodges and Harvey, 2003), and acid−base chemistry (McClary and Bretz, 2012). The most comprehensive diagnostic tool for acids and bases was developed by McClary and Bretz. These researchers capitalized on prior work when developing their ACID-I diagnostic (Cartrette and Mayo, 2011; McClary and Talanquer, 2011a, 2011b); ACID-I is a multi-tier multiple choice assessment designed to measure alternative conceptions regarding acid strength, and how tightly students hold onto these misconceptions. However, one major limitation of an assessment like this is that students are forced to choose one of the answers that is already on the page, and that may give instructors the illusion that their

المنارة للاستشارات
www.manaraa.com

students are harboring certain misconceptions when in fact they have only chosen those answers due to the forced-response format.

Open-ended (or constructed-response) assessment items lie between the extremes of interviews and multiple-choice assessments, and can provide a deeper understanding of student knowledge. Cooper *et al.* (2016) reported the development of a constructed-response item designed to measure use of acid−base models when explaining the outcome of a single-proton acid−base reaction, in which students were asked to describe both what was happening and why it was happening on a molecular level. The study found that student responses could be characterized as invoking a Brønsted-Lowry model or a Lewis model, and whether the response was mechanistic or causal. Mechanistic responses are surface-level, describing what is happening in the reaction, but not giving a reason for why it is happening. Causal responses include reasoning for why the reaction was occurring. The Cooper *et al.* study found that students who invoked a causal Lewis-based reasoning were more likely to draw mechanistic arrows for the reaction correctly, suggesting a higher-level understanding of acid−base theory. Constructed-response assessments are argued to provide deeper insight into thinking than forced-response assessments (Cooper et al., 2016; Birenbaum and Tatsuoka, 1987; Stowe and Cooper, 2017; Underwood et al., 2018); however, constructed-response items are not practical for quick scoring, especially when used in large enrollment courses. Constructed-response items require analysis that extends beyond the limitations of most online and in-class response systems (e.g. iClicker; Fies and Marshall, 2006). The utility of constructed-response items is consequently reliant on methods for analyzing open-ended responses efficiently and effectively.

3.4.3 Use of lexical analysis to understand student conceptions

Lexical analysis techniques provide a means for analyzing open-ended, essay-based responses in little time (Haudek *et al.*, 2011). For lexical analysis, a computer codes text-based responses for words and phrases. These codes are combined into categories and used to develop a predictive model that matches a

المنارة للاستشارات
www.manaraa.com

human-coded score; the predictive model and computer coding scheme can be applied to new text-based responses for verification and for use in classroom assessment. Lexical analysis has been used to construct a number of tools for evaluating a range of biology-related concepts (Ha *et al.*, 2011; Ha and Nehm, 2016; Haudek *et al.*, 2012; Nehm *et al.,* 2012; Nehm and Haertig, 2012; Prevost *et al.*, 2016; Weston *et al.*, 2015) and a limited range of statistics concepts (Kaplan *et al.*, 2014). There are 100+ constructed-response questions in the Automated Analysis of Constructed Response (AACR) library, a web-based repository of assessment responses built using lexical analysis (McCourt *et al.*, 2017). Instructors can submit student responses to the AACR website for analysis (https://create4stem.msu.edu/project/aacr/questions). There are a handful of items in the AACR library that address chemistry concepts; however, these items are limited to concepts that are central to the undergraduate biology curriculum (e.g., non-covalent interactions, acid/base strength, pH; Haudek *et al.*, 2012).

The process to develop a single constructed-response assessment item is lengthy (Urban-Lurain *et al.*, 2009); however, the end result is a tool that can effectively evaluate learning. Development begins with a large number of human-scored responses. Data are divided into a training set and a testing set. Developers build a coding scheme based on word and phrase use in line with human-scored responses; synonyms and commonly misspelled words are entered into the coding scheme. Codes are used to build a binary, ordinal, or multinomial logistic regression model based on the desired 'scoring' of the constructed-response. For example, a binary logistic regression model would be built if a correct (score = *1*) or incorrect (score = *0*) evaluation was desired. An ordinal logistic regression model would be built if a partial credit model was desired: *0* = incorrect, *1* = partially correct, *2* = completely correct. And, a multinomial logistic regression model would be built if a categorical score was desired: *1* = use of a plum pudding model to describe an atom, *2* = use of a Bohr model to describe an atom, *3* = use of a quantum mechanical model to describe an atom.

المنارة للاستشارات
www.manaraa.com

3.4.4 Summary of literature

Students struggle to understand acid−base models. The Lewis acid−base model, in particular, is key to explaining concepts and mechanisms central to the study of organic chemistry. Use of the Lewis acid−base model, however, in explaining acid−base reactions has not yet been shown to relate to performance on summative assessments. Constructed-response questions are illuminating formative assessments; however, grading and analysis constraints limit use, especially in large enrollment courses. Lexical analysis software can alleviate this constraint. A large database of constructed-response items has been developed for use in postsecondary biology courses. Except for a few cross-disciplinary concepts, the development of constructed-response assessment items scored with predictive logistic regression models have been limited in chemistry. Our study will consider the importance of assessing use of Lewis acid−base models in constructed-response assessment items and how lexical analysis can be used to analyze and grade an acid−base constructed-response item.

3.5 Research questions

Our study is guided by two research questions:

1. Do students who invoke the Lewis acid−base model when reasoning about a proton-transfer acid−base reaction perform statistically higher on acidity and basicity examination items than students who do not invoke a Lewis acid−base model?

2. Does lexical analysis lead a valid logistic regression model for predicting use of the Lewis acid−base model when reasoning about a proton-transfer acid−base reaction?

www.manaraa.com

Water and hydrobromic acid react to form hydronium and bromide anion.

A. **Describe in full detail *what* you think is happening on the molecular level for this reaction. Be sure to discuss the role of each reactant.**

B. **Using a molecular level explanation, please explain *why* this reaction occurs. Be sure to discuss why reactants form the products shown.**

C. **Please draw a reaction mechanism for how this reaction occurs including all curved arrows, lone pairs, and formal charges.**

Figure 3.2. Written homework assessment item (modified from Cooper *et al.*)

3.6 Methods

We conducted this work under application Pro#00028802, "Comprehensive evaluation of the University of South Florida's undergraduate and graduate chemistry curricula" as reviewed on December 13, 2016, by the University of South Florida's Institutional Review Board.

3.6.1 Research Question 1

Data were collected from two sections of the first-semester of a yearlong organic chemistry course (N=420 total students) taught by the same instructor at a large, research-intensive, public university in the Southeastern United States. The constructed-response item reported by Cooper *et al.* was modified for our study (see Figure 3.2 for item used in our study). There are two main modifications: First, the reaction was changed to water and hydrobromic acid. Second, the reaction was described in the prompt using words rather than Lewis structures; this was to encourage students to use chemical names when writing in sentence form, as well as to assess fluency in converting between chemical names and structures. The item was administered as a portion of a weekly written homework assignment and was graded for completion; it was administered one week prior to the term examination covering acids and bases. Most students (n = 410) completed the assignment. Six responses were excluded due to legibility and failing to complete all parts of the assessment item; the final sample included 404 responses.

المنارة للاستشارات
www.manaraa.com

A term examination was given one week after data from the constructed-response item were collected. The examination covered introductory concepts such as chemical structures, nomenclature, functional groups, acidity, polarity, and infrared spectroscopy. Thirty-two points on the examination were directly related to understanding and applying acid−base models, including predicting the products of acid−base reactions, determining which side of a reaction that is favored at equilibrium, and predicting relative acidity of a pair or series of compounds (see Appendix A for the acid−base examination items). Students had 90 minutes to complete the exam. The examinations were graded by the teaching staff, each of whom was assigned a specific set of item(s) on the examination to grade with a predetermined rubric. Grading was spot checked by the instructor of the course after all items had been scored and a total score was determined for the examination. All 404 students who completed the constructed-response item took the term examination. A Welch's T-test (Welch, 1947) was performed to compare mean scores on the examination acid−base items by whether the student did or did not invoke a Lewis acid−base model on the constructed-response item. Responses to the assessment were coded by use of an Arrhenius, Brønsted-Lowry, or Lewis acid−base model; responses were coded based on terminology and ideas associated with each model. Similar to data reported in the development of the item, students did not clearly differentiate their responses to the "how" and "why" components of the item (Cooper *et al.*, 2016); therefore, we too have combined responses to these two parts of the assessment for our analyses. Early attempts to force responses into use of a single model were counterproductive; therefore, each response was coded with any or all of the acid−base models. The causal coding scheme, as used for the original open-ended item (Cooper *et al*), placed student responses into a category that included use of a single acid−base model. Our student responses predominantly included mixed-models use; as such, we did not apply the coding scheme used by Cooper *et al.*

Responses were coded *Arrhenius* if the student wrote about how HBr dissociates in water, or if they made the claim that acid strength is based on how well a compound dissociates in water. Responses were

المنارة للاستشارات
www.manaraa.com

Table 3.1 Example responses to assessment item

Response	Arrhenius	Brønsted−Lowry	Lewis
"In this reaction, we are seeing a strong acid (HBr) getting completely ionized in water."	X		
"HBr acts as an acid and donates a proton to H_2O. In other words, H_2O acts as a base and receives a proton."		X	
"The bond between HCl is breaking and the electrons in the bond are going to Br as a lone pair."			X
"Hydrobromic acid is a strong acid due to its electronegativity and greater atomic size; it will dissociate almost completely. Since it is a strong acid, the forward reaction, and thus formation of a weak conjugate base is favored, the low bond strength of hydrobromic acid allows it to donate a proton to water, thereby forming hydronium."	X	X	
"Hydrobromic acid acts as the electrophile, accept electrons. Water=nucleophile. Also, the hydrobromic acid is a strong acid, so it dissociates completely in water."	X		X
"The water is the Brønsted−Lowry base (nucleophile) and the hydrobromic acid is the Brønsted−Lowry acid (electrophile). The nucleophile takes a proton from HBr to form hydronium (conjugate acid) and electrophile receives electrons from the HBr bond to form bromide anion which is the conjugate base."		X	X
"This reaction occurs to form the products as the acid dissociates when reaction with the base and during the proton transfer the Br− forms to become stable as it gains electrons and loses the hydrogen proton."	X	X	X

coded *Brønsted-Lowry* if the student wrote about the transfer of a proton or hydrogen. This includes the base "takes the hydrogen" or the acid "loses the hydrogen," or that a bond forms between the water and a hydrogen. Responses were coded as *Lewis* if the student mentioned the transfer of electrons or an electron pair, if the student wrote about electrons attacking, or if the student noted partial charges or electron density as a reason for why the reaction happened. Responses were not coded as *Lewis* for simply for mentioning the presence of electrons or lone pairs; an action verb or implied action was necessary. Examples of coded responses are provided in Table 3.1.

Author AJD originally coded all responses. Author JRR, then, independently coded 100 randomly selected responses (25%). Author AJD and JRR initially agreed on 76% of the items. After discussing disagreements, codes were changed for 18 responses; in the end there was 94% agreement. AJD then reevaluated the complete data set in light of the discussion and adjudication of disagreements.

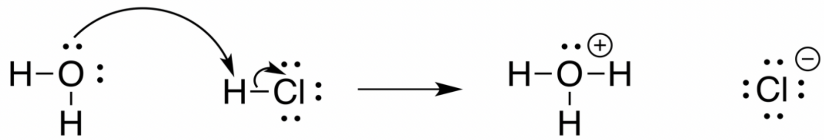
المنارة للاستشارات
www.manaraa.com

3.6.2 Research Question 2

Data were collected from five sections of the first-semester of a yearlong organic chemistry course taught by two instructors at a large, research-intensive, public university in the Southeastern United States. The constructed-response item used to answer Research Question 1 was modified to fit the new medium, which was a survey via Qualtrics rather than paper and pencil. Lewis structures and mechanistic arrows were added to the item prompt to reflect the focus of the item on what was occurring and why it was occurring portions of the assessment (see Figure 3.3); additionally, we did not have access to an online assessment system (such as BeSocratic) that allowed for graphical responses such to capture students drawing the reaction and reaction mechanism. While such tools are able to capture more drawing-based responses, grading still takes place by hand or through the use of very sensitive electronic grading. With our goal to address time constraints for assessment techniques used in large enrollment courses, more robust tools to grade hand-drawn pictures are necessary. In total, 752 responses were collected and human-coded by acid−base model use as described in the methods for Research Question 1. (Note: We have introduced the term *human-coded* into the description of methods, as to delineate computer-coding and predicted-codes germane to the lexical analysis methods used.) Responses used for addressing Research Question 2 were collected after a more summative assessment of acid−base understanding. Additionally, the 752 responses do include new responses from students who provided responses for Research Question 1. The purpose of Research Question 2 is to develop the predictive tool and not used in an experimental study; thus, the goal in collecting data for Research Question 2 is to obtain a wide array of responses rather than a snapshot of knowledge at a particular point in the course from which to evaluate an association with another metric. Students received extra credit toward their final examination grade (less than 1% of total final examination grade) for completing the assessment.

المنارة للاستشارات
www.manaraa.com

Consider the mechanism below for the acid-base reaction between water and hydrochloric acid to form hydronium ion and chloride ion.



A. Describe in full detail *what* you think is happening on the molecular level for this reaction. Be sure to discuss the role of each reactant.

B. Using a molecular level explanation, please explain *why* this reaction occurs. Be sure to discuss why the reactants form the products shown.

Figure 3.3. Online survey question given to students via Qualtrics

3.6.3 Lexical analysis and logistic regression model

Lexical analyses were conducted using SPSS Modeler Premium Text Analytics (version 18; herein referred to as SPSS Modeler).

Data were cleaned to remove extra spaces and HTML characters that may have been mistranslated when downloading the data from Qualtrics. Misspelled words were not corrected; misspellings were entered as synonyms for correctly spelled words when building the computer-coding scheme. Data were randomly divided into a training set (n = 534) and a testing set (n = 218), approximately a 70:30 split in the data (e.g., Ha *et al.*, 2016), which, based on our number of responses, allowed us to have a sufficient amount of responses to build the model while still leaving some left to test the model. The computer-coding scheme was built exclusively using the training set data. Common words and phrases were extracted from the data using built-in mechanisms within SPSS Modeler. Extracted words were assigned a type: For example, a type called *accept* was created to account for any word or phrase that meant something was accepting or was being accepted. Another example, a type called *chloride* was created to account for any word or phrase that referred to chloride (e.g., Cl–, chloride, chlorine ion). Types are used to build rules that recognize patterns in the text. For example, the computer would code a phrase as “chloride accepts” if a pattern of a chloride-type word was separated by a gap of three or less words followed by an accept-type word; thus,“the chloride

www.manaraa.com

was able to accept" would be coded as"chloride accepts". Individual codes are grouped into categories, i.e. words, phrases, and rule-based codes that have a homogenous meaning. For example, the"accept proton" category contains responses that invoke the acceptance of a proton, either through the pattern made from the types *accept* and *proton*, or from including both the word "accept" and the word "proton" in the response. A total of 28 categories were defined (See Appendix A for a complete list of categories and associated words, phrases, and rule-based codes.)

Categories were evaluated as to whether they predict the use of an Arrhenius, Brønsted-Lowry, or Lewis acid−base model. Categories associated with a single acid−base model were used as predictors in a binary logistic regression model to predict Lewis acid−base model use. There were not a sufficient number of responses coded as sole use of Arrhenius or Brønsted to build a predictive model for them; As we had found that use of a Lewis acid−base model significantly impacts student performance on an acid−base related assessment, it was decided that building a model to predict use vs. non-use of a Lewis acid−base model would be the most meaningful to organic chemistry instructors.

A binomial logistic regression model was built based on the computer-coding of the training set data. Logistic regression is a statistical method, similar to linear regression, that uses multiple variables to predict a dependent variable (Kleinbaum and Klein, 2010; Lewis, 2014; Prevost *et al.*, 2016). For our regression, the human-coded "Use" (*1*) and "Non-Use" (*0*) of a Lewis acid−base model is the dependent variable. Computer-coded categories are the independent variables. Regression coefficients (β) from the model are evaluated to determine the amount each lexical category contributes to the model logit. A logit is the natural log of P/(1-P), where "P" is the probability of a student's response being classified as "Use" of a Lewis acid−base model. The odds ratio for each category is determined by the exponential function of β , i.e. e^{β}; odds ratios are interpreted as the probability that a response will be coded as "Use" of a Lewis acid−base model when the category value is *1* and all other categories (i.e., independent variables) are equal to *0*.

www.manaraa.com

3.7 Results

3.7.1 Research Question 1: Impact of acid−base model used on examination scores

Our first goal was to establish if use of a Lewis acid−base model to explain a proton-transfer acid−base reaction (i.e., the formative assessment) was associated with higher performance on acid−base summative assessment items. These assessment items can be found in the Appendix A. Of the 404 usable formative assessment responses, 238 (59%) were human-coded as using a Lewis acid−base model in the student's explanation of what and why for the given reaction. 32 points of the examination included acid−base themes; A mean score of 21.6 (SD = 5.7) was observed for the acid−base items on the term examination. There was a significant difference in mean scores for those students who utilized a Lewis acid−base model on the formative assessment (M = 22.6, SD = 5.2) and those students who did not utilize a Lewis acid−base model on the formative assessment (M = 20.2, SD = 5.9); $t(362) = -4.33$, $p < 0.0001$, Cohen's $d = 0.42$ (small to medium; Cohen, 1988). The order of magnitude is classified according to Cohen's recommendations; however, when placed in the context of broader education research, the difference between the two groups' performance on the summative assessment items is comparable to studies shown to have high impact on learning and achievement (Hattie, 2008).

Given this finding, we recommend that instructors evaluate Lewis acid−base model use by their students and address observed deficiencies in learning through modified or remedial instruction. Formative assessment using this item, however, is limited by the time it takes to hand-score responses and prohibitively limited in utility for large enrollment courses. The assessment item was originally graded for completion when used in the course; the analysis reported herein took considerably longer than would have been appropriate for formative use of the assessment item. The development of an analysis tool to efficiently and

المنارة للاستشارات
www.manaraa.com

Table 3.2 Descriptive summary of categories by Lewis acid-base model use and results of binomial logistic regression

	Use (n=366)		Non-Use (n=168)				
Category	No.	%	No.	%	β	OR	p
Sharing electrons	48	13.1	1	0.6	3.18	24.15	0.009[a]
Partial charges	50	13.7	5	3.0	3.05	21.28	<0.001[a]
Bond electrons	172	47.0	11	6.5	2.72	15.21	<0.001[a]
Electron pairs	194	53.0	15	8.9	2.28	9.73	<0.001[a]
Attraction of hydrogen	68	18.6	7	4.2	2.00	7.40	0.001[a]
Nucleophile/electrophile	120	32.8	29	17.2	1.77	5.86	<0.001[a]
Accept electrons	100	23.3	10	6.0	1.08	2.93	0.036[a]
Donate electrons	95	26.0	9	5.4	0.68	1.96	0.201
Oxygen as an electron donor	26	7.1	1	0.6	0.58	1.78	0.656
Electronegativity	75	20.5	14	8.3	0.36	1.43	0.491
Donate protons	22	6.0	25	14.9	-0.27	0.76	0.597
Forming ions	19	5.2	12	7.1	-0.39	0.68	0.548
Accept protons	84	23.0	59	35.2	-0.51	0.60	0.116
Conjugate acid/base	18	4.9	22	13.1	-0.72	0.49	0.186
Acid strength	51	13.9	41	24.4	-1.25	0.29	0.005[a]
Hydrogen actions	61	16.7	24	14.3	-1.89	0.15	<0.001[a]
Hydrogen and not electrons	34	9.3	106	63.1	-2.08	0.13	<0.001[a]
Dissociation	33	9.0	41	24.4	-2.57	0.08	<0.001[a]
Regression constant					0.23		

[a] $p < 0.05$. $\chi^2(18) = 368.6$, $p < 0.0001$. OR = odds ratio

effectively score the item, either in the context of a classroom setting or through an online assessment system, would alleviate the aforementioned limitation.

3.7.2 Research Question 2: Development of a predictive model based on lexical analysis

Data from the training set (n = 534) were computer-coded as described in *Methods*. The number of responses in each category by "Use" and "Non-Use" of the Lewis acid−base model are reported in Table 3.2. In this table, "Use" represents a response being coded as positive for the use of a Lewis acid−base model and "Non-Use" represents a response being coded as negative for the use of a Lewis acid−base model. The results of a binomial logistic regression with human-coded Lewis acid−base use as the dependent variable and the 18 computer-coded categories as independent variables are also reported in Table 3.2. β represents the regression coefficient for each variable in the equation. The odds ratio is a measure of how much more or less likely a response is to be invoking a Lewis model if it is positive for a specific category. For example,

المنارة للاستشارات
www.manaraa.com

the category *sharing electrons* has an odds ratio of 24.15, which means that a response coded as positive for *sharing electrons* has 24.15 times higher odds of using a Lewis acid−base model than not invoking a Lewis acid−base model holding all other factors constant. Therefore, the larger the odds ratio value, the more likely that a response coded with that category is predicting a positive outcome (i.e., the Lewis acid−base model is used). The more negative the odds ratio value, the more likely that a response coded with that category is not predicting a positive outcome (i.e., the Lewis acid−base model is not used). The resultant binomial logistic regression model has a 0.87 prediction accuracy (i.e., the computer correctly predicted 87% of the human-codes); this level of prediction exceeds 0.70 accuracy level recommendation for predictive coding when using formative assessment data (e.g., Haudek *et al.*, 2012; Prevost *et al.*, 2016; Nehm *et al.*, 2012).

The resulting computer-coding scheme and predictive binomial logistic regression model were applied to the testing set data (n = 218). A 0.82 prediction accuracy level was obtained; this value exceeds recommended prediction accuracy levels for text analysis of formative assessment data. For human-coded "use" responses (n = 160), 88% were correctly predicted as "use;" for human-coded "non-use" responses (n = 58), 66% were correctly predicted as "non-use." Overall, predictions were 82% accurate.

3.8 Discussion and implications

Many have suggested that understanding of and ability to apply the Lewis acid−base model is critical for learning and achievement in organic chemistry coursework (Bhattacharyya, 2013; Cooper *et al.*, 2016; Nataro *et al.*, 2004; Stoyanovich *et al.*, 2015). From the global view of an instructor and organic chemist, the Lewis acid−base model is integral to thinking about the how and whys of reaction mechanisms such that the stated claim is warranted (Duis, 2011). However, there was a lack of empirical evidence in the literature to support the notion. Our results corroborate the claim and suggest that use of the Lewis acid−base

www.manaraa.com

model to explain a proton-transfer acid–base reaction on a formative assessment is associated with higher performance on summative items assessing acid–base knowledge. The results we present are narrow (i.e., a single formative assessment item associated with a bank for summative assessment items) and do not imply causality (many more studies are necessary for a causal claim to be made). However, these results do point to the importance of an instructor wanting to evaluate use of the Lewis acid–base model and the need for additional research on how and why students conceptualize acid–base reactions (and more broadly, all chemical reactions) using ideas from the Lewis acid–base model and other electrostatic and stochastic concepts. Our results also point to a need for studies that evaluate predictive factors and confounding influences on Lewis acid–base model use, including (1) prior instruction and learning before taking an organic chemistry course, (2) the relationship of model use with ability to conceptualize abstract ideas, and (3) the wording and presentation of the assessment item prompts.

For 62.1% of the responses used to answer our first research question and 72.1% of the responses used to answer our second research question, more than one acid–base model was used by the respondent in stating what occurred in the reaction and why it occurred. The Lewis acid–base model theoretically encompasses the Brønsted-Lowry model which in turn theoretically encompasses the Arrhenius model. While instructors and practicing chemists see clear delineations of the models and would thus be reticent to mixing models when explaining chemical reactions, a majority of students responding to the formative assessment item mixed the use of acid–base models. We can think of no occasion in the contexts of our research when students were told to choose only one model; in fact, we were resistant to providing students with such direction. We assert this lack of direction (or limitations) on how to respond to the assessment item allowed for a more accurate appraisal of student learning. However, use of multiple models when responding to the assessment item has implications for the utility of the predictive regression model built and evaluated in response to Research Question 2.

المنارة للاستشارات
www.manaraa.com

The results of the lexical analysis, including building a computer-coding scheme and evaluating a binomial logistic regression model, suggest that the technique is viable for use in predictive-scoring of formative uses of the proton-transfer acid–base assessment item (82% prediction accuracy with the testing set data). However, a super-majority of responses (>57% for data in both studies) were coded as "Use" of a Lewis acid–base model and additionally use of one or both of the other two models. This naturally occurring artifact of the data confounds the ability to identify clean predictors in a logistic regression model. We have reported the number of "Use" and "Non-Use" of Lewis acid–base model responses by computer-coded category to better understand the prevalence of ideas invoked by a respondent (see Table 3.2). Significant versus non-significant predictors in the logistic regression model should not be interpreted as concepts indicative or not indicative of Lewis acid–base model use. To a degree, the model has limited utility; for its intended purpose, the overall model is acceptable for providing predictive coding of responses as using or not using a Lewis acid–base model.

Recommendations for acid–base instruction based on empirical studies, scholarship of teaching and learning reports, and general discussions of techniques used are reported elsewhere; references are provided for those interested in considering how to improve acid–base instruction (Adcock, 2001; Barke *et al.*, 2009; Carlton, 1997; Drechsler and Schmidt, 2005; Drechsler and Van Driel, 2008; Erduran, 2003; Furió-Más *et al.*, 2005; Kala *et al.*, 2013; Kousathana *et al.*, 2005; Nakhleh and Krajcik, 1994; Sesen and Tarhan, 2011; Seyhan and Morgil, 2007; Shaffer, 2006; Stoyanovich *et al.*, 2015; Tarhan and Sesen, 2012). For our data and results, we have no foundation from which to make recommendations on best strategies for teaching acid–base theory, nor do we have a foundation from which to suggest remediation if students are found to not understand, misapply, or fail to use a Lewis acid–base model when explaining acid–base reactions. Additionally, we have no foundation for recommending when the assessment item is best given to inform instruction (e.g., before formal instruction as a pre-assessment, during or after instruction as a formative assessment). When collecting data for both our studies, the assessment item was given after

المنارة للاستشارات
www.manaraa.com

formal instruction on acid−base theory. Further work should consider the utility of the assessment item to provide meaningful data from which an instructor can modify and tailor instruction to address desired learning goals. *We recommend that the assessment item, computer-coding scheme and predictive logistic regression model be used as a formative assessment with little putative bearing on course grades as the prediction accuracy is not sufficient to be definitive in predicting human-coding.*

3.9 Limitations

In addition to the limitations presented in the *Discussion and implications* section (i.e., use of mixed models in responses, and using the assessment item in a formative context), two additional limitations should be considered: (1) the homogeneity of the sample population, and (2) availability of the software for computer-coding and predictive scoring.

First, data were collected from a single institution. While the distribution of responses suggests that we obtained data in which students used a single model and most combinations of the three models in their responses, these responses must be considered in the broader context of the instructor, course, and curriculum of the institution where the data were collected. We have already conceded that we were unable to code our data from a "causal" perspective as was originally done when the assessment item was developed and reported by Cooper *et al.* (2016); we speculate that this is a product of the instructors, course, and departmental setting in which the original work was done. The students sampled in our study were not exposed to the innovative curricula enacted by the authors of the original assessment item. This suggests that our computer-coding scheme and predictive logistic regression model should be further evaluated in the context of other contexts (e.g., institutions enacting the CLUE curriculum, utilizing active-learning strategies such as POGIL, enacting 1:2:1 course sequences, or curricula emphasizing science practice skills). While a

المنارة للاستشارات
www.manaraa.com

generalizable predictive model would be ideal, more work is necessary to identify the influence of various contexts on the predictive validity of the binomial logistic regression model we have reported.

Second, the computer-coding and evaluation of the binomial logistic regression model were conducted in SPSS Modeler Premium Text Analytics (version 18). It is nearly impossible to find an online quote of the software's price; author JRR was able to obtain a quote for limited use of the software after submitting an online request and speaking by phone with an IBM representative. Ultimately, we were able to obtain the software through a University of South Florida (our home institution) license. Software costs, thus, limit the availability and ease of dissemination of our research. (We do, however, invite and encourage readers to contact us regarding obtaining the files necessary to conduct analyses of their data using our coding scheme and predictive logistic regression.) There are other, more publicly accessible options including open-access text analytic software, that we are considering moving our work into; however, many of these options pale in comparison to the features and complex abilities of SPSS Modeler. Work in this area must obtain a delicate balance between high-quality analyses and the ability to widely disseminate the research.

3.10 Conclusion

Lewis acid−base model use is important for learning in organic chemistry courses. Our results corroborate a widely-held view that Lewis acid−base model understanding is necessary for achievement. The use of an open-ended, constructed-response assessment item designed to measure student understanding of acid−base models was limited in its utility for use in formative assessment by the time necessary to score the assessment. We presented a solution using lexical analysis that can efficiently and effectively code data collected from the assessment item; our prediction accuracies for training and testing set data were 87% and 82%, respectively. Further work should explore the predictive validity of our logistic regression model

المنارة للاستشارات
www.manaraa.com

in other instructional contexts (e.g., time of formative assessment, course curricula, use of active learning strategies).

3.11 Associated content

Examination items with acid−base content, categories included in the logistic regression, and categories not included in the logistic regression model can be found in Appendix A.

3.12 Acknowledgments

We would like to thank Luanna Prevost and Kelli Carter (University of South Florida) for introducing us to the SPSS Modeler software and providing troubleshooting advice. We would also like to thank the students who completed the assessment as part of their written homework and online.

3.13 References

Adcock, J.L., (2001), Teaching Bronsted-Lowry acid−base theory in a direct comprehensive way. *J. Chem. Educ.*, **78**, 1495-1496.

Arjoon, J.A., Xu, X., Lewis, J.E., (2013), Understanding the state of the art for measurement in chemistry education research: examining the psychometric evidence. *J. Chem. Educ.*, **90**, 536–545.

Barke, H.-D., Hazari, A., Yitbarek, S., (2009), Acid–Base Reactions, in: Misconceptions in Chemistry. Springer, Berlin, Heidelberg, pp. 173–206.

Bhattacharyya, G., (2013), From source to sink: mechanistic reasoning using the electron-pushing formalism. *J. Chem. Educ.*, **90**, 1282–1289.

Bhattacharyya, G., (2006), Practitioner development in organic chemistry: how graduate students conceptualize organic acids. *Chem. Educ. Res. Pract.*, **7**, 240–247.

Black, P., Wiliam, D., (1998), Inside the black box: raising standards through classroom assessment. *Phi Delta Kappan*, **80**, 139–44.

Bretz, S.L., Linenberger, K.J., (2012), Development of the enzyme–substrate interactions concept inventory. *Biochem. Mol. Biol. Educ.*, **40**, 229–233.

Bretz, S.L., McClary, L., (2015), Students' understandings of acid strength: how meaningful is reliability when measuring alternative conceptions? *J. Chem. Educ.*, **92**, 212–219.

www.manaraa.com

Caldwell, J.E., (2007), Clickers in the large classroom: current research and best-practice tips. *CBE-Life Sci. Educ.*, **6**, 9–20.
Carlton, T.S., (1997), Why and how to teach acid−base reactions without equilibrium. *J. Chem. Educ.*, **74**, 939-941.
Cartrette, D.P., Mayo, P.M., (2011), Students' understanding of acids/bases in organic chemistry contexts. *Chem. Educ. Res. Pract.*, **12**, 29–39.
Christian, B.N., Yezierski, E.J., (2012), Development and validation of an instrument to measure student knowledge gains for chemical and physical change for grades 6–8. *Chem. Educ. Res. Pract.*, **13**, 384–393.
Cohen, J., (1988), Statistical power analysis for the behavioral sciences. L. Erlbaum Associates, Hillsdale, N.J.
Cooper, M.M., Kouyoumdjian, H., Underwood, S.M., (2016), Investigating students' reasoning about acid–base reactions. *J. Chem. Educ.*, **93**, 1703–1712.
Dicks, A.P., Lautens, M., Koroluk, K.J., Skonieczny, S., (2012), Undergraduate oral examinations in a university organic chemistry curriculum. *J. Chem. Educ.*, **89**, 1506–1510.
Domin, D., Bodner, G., (2012), Using students' representations constructed during problem solving to infer conceptual understanding. *J. Chem. Educ.*, **89**, 837–843.
Drechsler, M., Schmidt, H.-J., (2005), Textbooks' and teachers' understanding of acid−base models used in chemistry teaching. *Chem. Educ. Res. Pract.*, **6**, 19–35.
Drechsler, M., Van Driel, J., (2008), Experienced teachers' pedagogical content knowledge of teaching acid–base chemistry. *Res. Sci. Educ.*, **38**, 611–631.
Duis, J.M., (2011), Organic chemistry educators' perspectives on fundamental concepts and misconceptions: An exploratory study. *J. Chem. Educ.*, **88**, 346–350.
Erduran, S., (2003), Examining the mismatch between pupil and teacher knowledge in acid−base chemistry. *Sch. Sci. Rev.*, **84/308**, 81–87.
Fies, C., Marshall, J., (2006), Classroom response systems: a review of the literature. *J. Sci. Educ. Technol.*, **15**, 101–109.
Friesen, J.B., (2008), Saying what you mean: teaching mechanisms in organic chemistry. *J. Chem. Educ.*, **85**, 1515-1518.
Furió-Más, C., Calatayud, M.L., Guisasola, J., Furió-Gómez, C., (2005), How are the concepts and theories of acid–base reactions presented? Chemistry in textbooks and as presented by teachers. *Int. J. Sci. Educ.*, **27**, 1337–1358.
Gebru, M.T., Phelps, A.J., Wulfsberg, G., (2012), Effect of clickers "versus" online homework on students' long-term retention of general chemistry course material. *Chem. Educ. Res. Pract.*, **13**, 325–329.
Ha, M., Nehm, R.H., (2016), The impact of misspelled words on automated computer scoring: a case study of scientific explanations. *J. Sci. Educ. Technol.*, **25**, 358–374.
Ha, M., Nehm, R.H., Urban-Lurain, M., Merrill, J.E., (2011), Applying computerized-scoring models of written biological explanations across courses and colleges: Prospects and limitations. *CBE Life Sci. Educ.*, **10**, 379–393.
Hattie, J., (2008), Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement. Routledge.
Haudek, K.C., Kaplan, J.J., Knight, J., Long, T., Merrill, J., Munn, A., Nehm, R., Smith, M., Urban-Lurain, M., (2011), Harnessing technology to improve formative assessment of student conceptions in STEM: forging a national network. *CBE-Life Sci. Educ.*, **10**, 149–155.
Haudek, K.C., Prevost, L.B., Moscarella, R.A., Merrill, J., Urban-Lurain, M., (2012), What are they thinking? Automated analysis of student writing about acid–base chemistry in introductory biology. *CBE-Life Sci. Educ.*, **11**, 283–293.

المنارة للاستشارات
www.manaraa.com

Henderleiter, J., Smart, R., Anderson, J., Elian, O., (2001), How do organic chemistry students understand and apply hydrogen bonding? *J. Chem. Educ.*, **78**, 1126.

Herron, F.Y., (1953), Models illustrating the Lewis theory of acids and bases. *J. Chem. Educ.*, **30**, 199–201.

Hodges, L.C., Harvey, L.C., (2003), Evaluation of student learning in organic chemistry using the SOLO taxonomy. *J. Chem. Educ.*, **80**, 785.

Kala, N., Yaman, F., Ayas, A., (2013), The effectiveness of predict-observe-explain technique in probing students' understanding about acid−base chemistry: a case for the concepts of pH, pOH, and strength. *Int. J. Sci. Math. Educ.*, **11**, 555–574.

Kaplan, J.J., Haudek, K.C., Ha, M., Rogness, N., Fisher, D.G., (2014), Using lexical analysis software to assess student writing in statistics - eScholarship. *Technol. Innov. Stat. Educ.*, **8**.

King, D.B., (2011), Using clickers To identify the muddiest points in large chemistry classes. *J. Chem. Educ.*, **88**, 1485–1488.

Kleinbaum, D.G., Klein, M., (2010), Logistic Regression, Statistics for Biology and Health. Springer New York, New York, NY.

Kousathana, M., Demerouti, M., Tsaparlis, G., (2005), Instructional misconceptions in acid−base equilibria: an analysis from a history and philosophy of science perspective. *Sci. Educ.*, **14**, 173–193.

Kristine, F.J., (1985), Developing study skills in the context of the general chemistry course: The prelecture assignment. *J. Chem. Educ.*, **62**, 509-510.

Lewis, S.E., (2014), An Introduction to Nonparametric Statistics in Chemistry Education Research, in: Tools of Chemistry Education Research, ACS Symposium Series. American Chemical Society, pp. 115–133.

Luder, W.F., (1948), Contemporary acid−base theory. *J. Chem. Educ.*, **25**, 555–558.

MacArthur, J.R., Jones, L.L., (2008), A review of literature reports of clickers applicable to college chemistry classrooms. *Chem. Educ. Res. Pract.*, **9**, 187–195.

McClary, L., Bretz, S.L., (2012), Development and assessment of a diagnostic tool to identify organic chemistry students' alternative conceptions related to acid strength. *Int. J. Sci. Educ.*, **34**, 2317–2341.

McClary, L., Talanquer, V., (2011a), College chemistry students' mental models of acids and acid strength. *J. Res. Sci. Teach.*, **48**, 396–413.

McClary, L., Talanquer, V., (2011b), Heuristic reasoning in chemistry: making decisions about acid strength. *Int. J. Sci. Educ.*, **33**, 1433–1454.

McCourt, J.S., Andrews, T.C., Knight, J.K., Merrill, J.E., Nehm, R.H., Pelletreau, K.N., Prevost, L.B., Smith, M.K., Urban-Lurain, M., Lemons, P.P., (2017), What motivates biology instructors to engage and persist in teaching professional development? *CBE-Life Sci. Educ.*, **16**, ar54.

Menucha Birenbaum, Kikumi K. Tatsuoka, (1987), Open-ended versus multiple-choice response formats—it does make a difference for diagnostic purposes. *Appl. Psychol. Meas.*, **11**, 385–395.

Nakhleh, M.B., Krajcik, J.S., (1994), Influence of levels of information as presented by different technologies on students' understanding of acid, base, and pH concepts. *J. Res. Sci. Teach.*, **31**, 1077–1096.

Nataro, C., Ferguson, M.A., Bocage, K.M., Hess, B.J., Ross, V.J., Swarr, D.T., (2004), Lewis acid−base, molecular modeling, and isotopic labeling in a sophomore inorganic chemistry laboratory. *J. Chem. Educ.*, **81**, 722–724.

National Research Council, (2001), Knowing What Students Know: The Science and Design of Educational Assessment. The National Academies Press, Washington, DC.

Nehm, R.H., Ha, M., Mayfield, E., (2012), Transforming biology assessment with machine learning: automated scoring of written evolutionary explanations. *J. Sci. Educ. Technol.*, **21**, 183–196.

Nehm, R.H., Haertig, H., (2012), Human vs. computer diagnosis of students' natural selection knowledge: testing the efficacy of text analytic software. *J. Sci. Educ. Technol.*, **21**, 56–73.

Paik, S.-H., (2015), Understanding the relationship among Arrhenius, Brønsted–Lowry, and Lewis theories. *J. Chem. Educ.*, **92**, 1484–1489.

المنارة للاستشارات
www.manaraa.com

Prevost, L.B., Smith, M.K., Knight, J.K., (2016), Using student writing and lexical analysis to reveal student thinking about the role of stop codons in the central dogma. *CBE-Life Sci. Educ.*, **15**, ar65: 1-13.

Roecker, L., (2007), Using oral examination as a technique to assess student understanding and teaching effectiveness. *J. Chem. Educ.*, **84**, 1663.

Rushton, G.T., Hardy, R.C., Gwaltney, K.P., Lewis, S.E., (2008), Alternative conceptions of organic chemistry topics among fourth year chemistry students. *Chem. Educ. Res. Pract.*, **9**, 122–130.

Seery, M., (2012), Jump-starting lectures. *Educ. Chem.*, **49**, 22-25.

Seery, M.K., Donnelly, R., (2012), The implementation of pre-lecture resources to reduce in-class cognitive load: A case study for higher education chemistry. *Br. J. Educ. Technol.*, **43**, 667–677.

Sesen, B.A., Tarhan, L., (2011), Active-learning versus teacher-centered instruction for learning acids and bases. *Res. Sci. Technol. Educ.*, **29**, 205–226.

Seyhan, H.G., Morgil, I., (2007), The effect of 5E learning model on teaching of acid−base topic in chemistry education/Los efectos del modelo 5E de aprendizaje en la enseñanza del tema “Acidos-bases” en química. *J. Sci. Educ.*, **8**, 120.

Shaffer, A.A., (2006), Let us give Lewis acid−base theory the priority it deserves. *J. Chem. Educ.*, **83**, 1746–1749.

Stowe, R., Cooper, M., (2017), Practicing what we preach: assessing “critical thinking” in organic chemistry. *J. Chem. Educ.*, **94**, 1852–1859.

Stoyanovich, C., Gandhi, A., Flynn, A.B., (2015), Acid–base learning outcomes for students in an introductory organic chemistry course. *J. Chem. Educ.*, **92**, 220–229.

Taagepera, M., Noori, S., (2000), Mapping students’ thinking patterns in learning organic chemistry by the use of Knowledge Space Theory. *J. Chem. Educ.*, **77**, 1224–1229.

Tarhan, L., Sesen, B.A., (2012), Jigsaw cooperative learning: acid–base theories. *Chem. Educ. Res. Pract.*, **13**, 307–313.

Underwood, S.M., Posey, L.A., Herrington, D.G., Carmel, J.H., Cooper, M.M., (2018), Adapting Assessment Tasks To Support Three-Dimensional Learning. *J. Chem. Educ.*, **95**, 207–217.

Urban-Lurain, M., Moscarella, R.A., Haudek, K.C., Giese, E., Sibley, D.F., Merrill, J.E., (2009), Beyond multiple choice exams: Using computerized lexical analysis to understand students’ conceptual reasoning in STEM disciplines. Presented at the Proceedings - Frontiers in Education Conference, FIE.

Welch, B.L., (1947), The generalization of “student’s” problem when several different population variances are involved. *Biometrika*, **34**, 28–35.

Weston, M., Haudek, K.C., Prevost, L., Urban-Lurain, M., Merrill, J., (2015), Examining the impact of question surface features on students’ answers to constructed-response questions on photosynthesis. *CBE-Life Sci. Educ.*, **14**, ar19: 1-12.

المنارة للاستشارات
www.manaraa.com

Chapter 4

Development and evaluation of a Lewis acid−base tutorial for use in postsecondary organic chemistry courses

4.1 Note to reader

This chapter is a published manuscript in the Canadian Journal of Chemistry. Republished with permission of Canadian Science Publishing, from Development and evaluation of a Lewis acid−base tutorial for use in postsecondary organic chemistry courses, Dood, Amber J.; Cruz-Ramírez de Arellano, Daniel; Fields, Kimberly B.; Raker, Jeffrey R., 97, 6, 2019; permission conveyed through Copyright Clearance Center, Inc.

Further permissions information can be found in Appendix C.

This work was published with co-authors. Dr. Daniel Cruz-Ramírez de Arellano is an instructor of organic chemistry at the University of South Florida and provided feedback about tutorial activities as well as allowed for data collection in his classes. Dr. Kimberly B. Fields is an instructor of organic chemistry at the University of South Florida who allowed for data collection in her classes. Dr. Jeffrey R. Raker is the principal investigator for this project.

www.manaraa.com

4.2 Abstract

A well-developed understanding of the Lewis acid−base model is highly important for the understanding of organic chemistry. As such, students should receive instruction and be assessed on use of the model. Online tutorials and constructed-response items provide a means for confirming that students have a well-developed conceptualization of the Lewis acid−base model. In a prior study, a predictive logistic regression model was presented that can be used with constructed-response assessment items to determine use of a Lewis acid−base model in written responses. In this study, we use that predictive model to evaluate the effectiveness of a tutorial designed to promote meaningful understanding of the Lewis acid−base model in three different instructional contexts: first-semester organic chemistry students before summative assessment, first-semester organic chemistry students after summative assessment, and second-semester organic chemistry students. Additionally, we evaluated the learning gains of one set of first-semester students after a three-week time delay. McNemar's test results suggest that the tutorial had a net positive impact in all three instructional contexts, with the most significant impact observed with the second-semester students. This work has implications for further development of literature-based tutorials to promote meaningful understanding of organic chemistry reaction mechanisms assessed by constructed-response items.

4.3 Introduction

The Lewis acid−base model is the most useful acid−base model for explaining reactivity in organic chemistry (Bhattacharyya, 2013; Cooper *et al.*, 2016; Herron, 1953; Luder, 1948; Nataro *et al.*, 2004; Shaffer, 2006). It has been shown that the understanding and application of the Lewis model by students is associated with higher success on organic chemistry examination assessments (Dood *et al.*, 2018). The development and evaluation of learning tools that support understanding of the Lewis acid−base model has

www.manaraa.com

potential to increase understanding of many concepts in organic chemistry, including reactions and reaction mechanisms. Such tools should address prior knowledge and assess knowledge of Lewis acids and bases, including targeting of known deficiencies reported in the literature (Bretz and McClary, 2015; Cartrette and Mayo, 2011; McClary and Bretz, 2012; McClary and Talanquer, 2011a, 2011b; Paik, 2015; Tümay, 2016). One way to formatively assess acid−base model understanding is through constructed-response items (Cooper *et al.*, 2016; Dood *et al.*, 2018). It has been shown that a predictive binomial logistic regression model, based on text analyses, can determine whether students invoke Lewis acid−base model concepts for a single acid−base proton-transfer reaction constructed-response item (Dood *et al.*, 2018). The results from that study corroborate the presence of well-known misconceptions of acids, bases, and acid−base models (Dood *et al.*, 2018). While implications have been offered in the literature for remediating deficiencies in understanding the Lewis acid−base model, there have been few accounts of the development and evaluation of research-based tools that promote model use when explaining chemical phenomena. Herein, we report on the development of a tutorial, informed by research, designed to promote use of the Lewis acid−base model in explaining the mechanism of a single proton-transfer reaction; we utilize our previously reported predictive model to analyze written explanations before and after the tutorial (Dood *et al.*, 2018). Statistical analyses suggest that the tutorial is effective in promoting understanding and use of the Lewis acid−base model.

4.4 Research Question

Our study was guided by one principal research question:

What is the impact of a targeted intervention to construct Lewis acid−base model concepts on student responses to an acid−base proton transfer reaction question about what happens in the reaction mechanism and why it happens?

المنارة للاستشارات
www.manaraa.com

4.5 Student understanding of acid−base models

Several studies have considered student understanding of acid−base models in chemistry (Bretz and McClary, 2015; Cartrette and Mayo, 2011; Cooper *et al.*, 2016; Demircioglu *et al.*, 2005; Erduran, 2003; Haudek *et al.*, 2012; Kousathana et al., 2005; Lin and Chiu, 2010; McClary and Bretz, 2012; McClary and Talanquer, 2011a, 2011b; Nakhleh and Krajcik, 1994; Sesen and Tarhan, 2011; Stoyanovich *et al.*, 2015). Many focused on the Arrhenius and Brønsted-Lowry acid−base models, i.e. models featured within secondary schools' chemistry courses and general chemistry courses at postsecondary institutions (Shaffer, 2006).

The Lewis model is a component of most secondary- and postsecondary-level chemistry courses; however, it is often not the primary focus (Shaffer, 2006). Students thus come to organic chemistry courses more comfortable with other models, such as the Brønsted-Lowry model, while being confronted with situations that require an understanding of the Lewis model. Model confusion occurs as students attempt to apply the Brønsted-Lowry model to contexts and problems where it is not appropriate. Students hold on to misconceptions related to their lack of understanding of a Lewis acid−base model while attempting to make meaning of new contexts and problems (Cartrette and Mayo, 2011). For example, Cartrette and Mayo found that organic chemistry students attempted to force concepts associated with the Lewis model (i.e., nucleophilicity and electrophilicity) into their existing understanding of acids and bases via the Brønsted-Lowry model. A student in the study by Cartrette and Mayo rationalized that boron trifluoride (BF_3) was a Lewis acid by explaining that it could not accept a proton and therefore could not be a Brønsted base. Other participants in the study incorrectly identified Lewis acids and bases by attempting to fit them to a Brønsted-Lowry model when a Lewis model was required. Another participant in the study was unable to make connections between acidity and basicity and nucleophilicity and electrophilicity, and, when prompted, claimed that a base "would be electrophilic because it's going for the hydrogen atom". All participants in

المنارة للاستشارات
www.manaraa.com

the study were able to correctly give the Brønsted-Lowry definition and come up with examples of Brønsted-Lowry acids and bases; less than half were able to do the same for the Lewis model. Students' heavy reliance on the Brønsted-Lowry model may contribute to their inability to make connections between Lewis acids and bases and nucleophiles and electrophiles in later courses (e.g., organic chemistry or inorganic chemistry).

Another study found student difficulties in understanding acid−base reactions that cannot be explained using the Brønsted-Lowry model (Tarhan and Sesen, 2012). Tarhan and Sesen found that, when considering the acid−base reaction between BF_3 and NH_3, 40% of students could not explain the reaction using the required Lewis model. Tarhan and Sesen found a persistent belief among students that there is no electron transfer between NH_3 and BF_3 molecules.

Taagepera and Noori (2000) identified several common misconceptions among general chemistry and organic chemistry students; of importance was the inability of students to recognize reaction types such as simple proton-transfer reactions. The researchers noted that students who were better at explaining electron densities were also better at recognizing simple proton-transfer reactions. While a simple proton-transfer reaction can be described using the Brønsted-Lowry model, the researchers indicated that experts base their understanding of chemical properties on electron densities, a concept which more appropriately relates to the Lewis model. Taagepera and Noori suggested that unless students are specifically taught to make connections between chemical properties and electron densities, students will not develop an appropriate expert level of understanding. The study also found that when instructors repeatedly emphasized electron density across concepts, rather than just explaining it once and casually mentioning it thereafter, student understanding became more similar to that of experts. Taagepera and Noori recommended making deliberate connections across concepts to allow students to fit new information into their existing knowledge of basic principles; instructors making these connections explicit allows students to develop a more expert-level of understanding.

المنارة للاستشارات
www.manaraa.com

Cooper *et al.* articulated that while the Brønsted-Lowry model explains the "what" of the reaction, the Lewis model provides the "how" (Cooper *et al.*, 2016). The Brønsted-Lowry model describes the proton transfer; the rationalization for how the reaction is occurring comes from the Lewis model, that is, a proton transfer happens due to electron pairs involved in breaking and forming bonds. An explanation of why a reaction occurs can also be constructed using the Lewis model. For example, reasoning behind reactions occurring such as the attraction of electrons to areas of low electron density and the interactions of partial charges are encompassed by the Lewis acid−base model. Cooper argued that students must understand the "why" in order to develop a deeper understanding of chemical phenomena (2015). Cooper *et al.* (2016) found that students who provided explanations that invoked the Lewis acid−base model rather than the Brønsted-Lowry model, regardless of whether or not the explanation described why the reaction occurred, were more likely to come up with correct reaction mechanisms; again emphasizing the importance of a well-developed understanding of the Lewis model for understanding organic chemistry. Students who used causal reasoning in their explanations (i.e., stated why the reaction occurs) were more likely to produce the correct reaction arrows than those who did not regardless of acid−base model used, but the importance of using the Lewis acid−base model over the Brønsted-Lowry model was more pronounced in terms of correlating with the correct production of mechanistic arrows. As the Lewis model emphasizes electrons, the argument from Cooper *et al.* aligns with the conjecture that expert level reasoning about proton-transfer reactions involves a discussion of electron density (Taagepera and Noori, 2000); a novice can explain "what" using the Brønsted-Lowry model while an expert can explain "what", "how", and "why" using the Lewis model.

Further studies on the understanding of organic chemistry mechanisms have supported this emphasis on electron density and the Lewis model (Bhattacharyya, 2013; Bhattacharyya and Harris, 2018; N. Grove *et al.*, 2012). The Lewis model closely relates to the concepts of electron density and nucleophilicity and electrophilicity, of which understanding has been found to correlate with success in solving organic chemistry mechanisms (Cooper *et al.*, 2016). In a study of student descriptions of organic mechanisms, Bhattacharyya

www.manaraa.com

and Harris (2018) found that students who were able to propose mechanisms for unknown reactions were primarily successful because they were able to identify specific atoms in the reactants and products that were the sources and sinks of electrons, i.e., the nucleophile and the electrophile.

Students are confused by amphoteric properties, finding it difficult to understand that water can act as both an acid and a base. For example, one student interviewed in a study by Drechsler and Van Driel (2008) said, “It is pretty tricky, because, if you look at water, it is both an acid and a base”. Other studies found that students were uncomfortable classifying water as an acid or as a base, possibly due to dependence on the Arrhenius acid−base model (Drechsler and Schmidt, 2005; Schmidt and Volke, 2003). Additionally, Ültay and Calik (2016) found that pre-service chemistry teachers held the misconception that water cannot act as an acid or a base because it acts as a solvent. Schmidt (1997) identified several misconceptions about acids and bases. One misconception was that acids must hold a positive charge and that bases must hold a negative charge; thus the misconception follows that neutral compounds, e.g., water, are neither an acid nor a base. Though the amphoteric property of water can be described using the Brønsted-Lowry model, not all amphoteric substances are proton donors. Therefore, understanding the amphoteric property of water through a Lewis model lens is important for understanding the acidity and basicity of a wider variety of substances. Based on the corpus of research literature regarding acidity and basicity, the development of any learning tools and assessments targeting acid−base understanding should consider helping students identify the scope of each model, including what contexts the model can and should be applied to when solving problems. Students should understand that the Brønsted-Lowry acid−base model is not applicable to most situations in organic chemistry (Cartrette and Mayo, 2011; Tarhan and Sesen, 2012). The Lewis acid−base model, in contrast, has a broader applicability; such situations include reactions that involve the transfer of electrons but not the transfer of protons. Even though simple proton-transfer reactions can be described using the Brønsted-Lowry model, students who describe such reactions using electron density have more expert-like reasoning (Taagepera and Noori, 2000). Understanding electron density and nucleophilicity

المنارة للاستشارات
www.manaraa.com

and electrophilicity have shown to be associated with greater success in solving organic chemistry reaction mechanisms (Bhattacharyya and Harris, 2018; Cooper *et al.*, 2016; Grove *et al.*, 2012). Additionally, understanding water's ability to act as both an acid and a base can improve student's overall understanding of Lewis acids and bases (Drechsler and Schmidt, 2005; Drechsler and Van Driel, 2008; Garnett *et al.*, 1995; Schmidt and Volke, 2003; Ültay and Calik, 2016).

4.6 Online tutorials

Meaningful learning can be promoted through research-based active learning techniques such as online tutorials (Allen, 1995). Individualized and adaptive learning in online environments have been shown to lead to higher achievement (Arasasingham *et al.*, 2011, 2005; Freasier *et al.*, 2003; He *et al.*, 2012; Korkmaz and Harwood, 2004; Littlejohn *et al.*, 2002; Sesen and Tarhan, 2011). One study implementing web-based instruction found that students using the online system performed better on assessments than those using textbook-based study strategies, and those students emerged from the course as better conceptual problem-solvers (Arasasingham *et al.*, 2005). In an introductory science course, Arasasingham *et al.* found that an online system emphasizing known student difficulties increased exam performance. In a study by Freasier *et al.* (2003) in a general chemistry course using online tutorials, voluntary completion of extra quizzes embedded in the tutorials was positively correlated with higher course grades. A supermajority (94%) of the students in that course felt that the online tutorials were helpful. Freasier *et al.* argued that the observed participation was the result of increased motivation for learning due to the online tutorials. In an analytical chemistry course, He *et al.* (2012) identified specific student needs based on difficulty in homework problems and created 10-15 minute online video tutorials for each specifically identified learning goal. He *et al.* found that online tutorials had a positive impact on student mastery of the subjects covered, particularly for average-performing and low-performing students. Littlejohn *et al.* (2002) designed tutorials for a

المنارة للاستشارات
www.manaraa.com

carbohydrate chemistry course. Students interviewed in an evaluation of the tutorials said that the tutorials were useful as a supplement to traditional lectures and would recommend their use to fellow undergraduates. Korkmaz and Harwood (2004) developed an interactive tutorial for learning molecular symmetry. Key aspects of the design were the succinct nature of the text, feedback provided, and inclusion of summative assessment. Overall, students found the tutorial valuable.

Online tutorials have been used in postsecondary organic chemistry courses (O'Sullivan and Hargaden, 2014; Richards-Babb *et al.*, 2015). Richards-Babb *et al.* (2015) implemented online, targeted assessments as a formative tool to increase learning in an organic chemistry course; the authors found that their students felt the tool was useful. Additionally, performance on the assessments correlated with the students' final grades in the course. O'Sullivan and Hargaden (2014) developed tutorials in their organic chemistry course that prompted students to draw chemical structures (i.e., predict the products and synthesis problems) and the system then provided targeted feedback on their structures. Overall, students agreed that the tutorials were a "beneficial learning experience".

Important features of the tutorials described above include emphasizing known student difficulties, formative assessment within the tutorial, conciseness, and feedback. The type and frequency of feedback in online tutorials is key to their effectiveness. Dempsey (1993) described feedback as "any information that follows a response and allows a student to evaluate the adequacy of the response itself". Feedback has shown a positive impact on retention and performance in various subjects (Jaehnig and Miller, 2007; Kulik and Kulik, 1988; Schimmel, 1983). Dempsey described the timing of feedback and the intensity of feedback. The feedback can be immediate (i.e., at the time of instruction or testing) or delayed. Kulik and Kulik (1988) published a meta-analysis on the timing of feedback and concluded that immediate feedback is more effective than delayed feedback. The intensity of feedback can vary; levels of corrective feedback include no feedback, simple verification feedback, correct response feedback, elaborated feedback, and try-again (or answer until correct) feedback (Dempsey, 1993). Simple verification feedback tells the student

المنارة للاستشارات
www.manaraa.com

whether their answer to a question is correct or incorrect. Correct response feedback tells the student the correct answer. Elaborated feedback gives the student a more detailed explanation regarding their answer, typically providing reasoning for why the student's response is correct or incorrect. Try-again feedback gives the learner simple feedback that the answer is incorrect and allows the question to be answered again. In a review, Jaehnig and Miller (2007) concluded that any feedback that allows the student to know the correct answer can be effective, with elaborated feedback being the most effective type of feedback.

4.7 Lexical analysis of constructed-response items as a formative assessment tool

Cooper *et al.* have argued that open-ended, constructed-response items are necessary to adequately assess student understanding of the Lewis acid−base model. However, such assessment items require time and personnel to "grade". Lexical analysis has shown promise for evaluating constructed-response items through computer scoring (Ha *et al.*, 2011; Ha and Nehm, 2016; Haudek *et al.*, 2012; Nehm and Haertig, 2012; Prevost *et al.*, 2016; Weston *et al.*, 2015). Over 100 items are available in the Automated Analysis of Constructed Response (AACR) library (https://create4stem.msu.edu/project/aacr/questions); most items in the library, though, target biology concepts (Haudek *et al.*, 2012). We remedied the burden of assessing student understanding of the Lewis acid−base model by capitalizing on the AACR project's lexical analysis methodology; we developed a predictive model for scoring written explanations of an acid−base proton-transfer reaction with 86% accuracy for "use" or "non-use" of the Lewis acid−base model in the response (Dood *et al.*, 2018). A key implication of the AACR work, and our own, is the use of the constructed-response items and computer-based scoring models as formative assessments; in theory, an instructor could give the item before, during, or after instruction and use the results to modify instruction, homework assignments, etc. Despite this recurring implication pronouncement, including our own pronouncement (Dood *et al.*, 2018), the use of these tools in classroom or online tutorial environments to increase understanding

المنارة للاستشارات
www.manaraa.com

has not been evaluated. A key opportunity for computer-based scoring of constructed-response items is their integration into online tutorial systems as pre/post measures of student understanding and as means for providing individualized tutorial components to address deficiencies in student understanding.

4.8 Summary of literature

Students struggle to understand acid−base models, particularly the Lewis acid−base model (Cartrette and Mayo, 2011). Notably, students struggle to relate the role of electrons in the forming and breaking of bonds in acid−base reactions. Online tutorials, particularly those targeting known difficulties and misconceptions, have shown positive impacts on student learning and garner positive feedback from the students who use the tools. A research-based method to determine student acid−base model use through formative assessment using constructed-response questions could be used to evaluate learning. We previously demonstrated that a logistic regression model can predict "use" or "non-use" of the Lewis acid−base model in response to a constructed-response prompt for a proton-transfer reaction (Dood *et al.*, 2018). Evaluation of the effectiveness of using these constructed-response items in classrooms or via online tutorials is absent from the literature. As such, we report the impact of an online tutorial grounded in a research-based understanding of common misconceptions of the Lewis acid−base model; we used an expanded version (i.e., accommodating multiple leaving groups) of our reported predictive model for scoring pre/post evaluations of student use of the Lewis acid−base model when responding to a constructed-response item.

المنارة للاستشارات
www.manaraa.com

4.9 Methods

This work was conducted under ethics application Pro#00028802, "Comprehensive evaluation of the University of South Florida's undergraduate and graduate chemistry curricula", as reviewed on December 13, 2016 by the Chair of the University of South Florida Institutional Review Board.

Herein, we describe the development of a Lewis acid−base tutorial, the expansion of our predictive logistic regression model to evaluate responses to a cloned constructed-response item, and the statistical test (i.e., McNemar test; McNemar, 1947; Cleophas and Zwinderman, 2016; Sheskin, 2011) for evaluating the effectiveness of the tutorial.

4.9.1 Development of the targeted Lewis acid−base tutorial

A tutorial was developed to promote understanding and use of the Lewis acid−base model, i.e., the "how" of the reaction, when explaining a proton-transfer reaction. The online tutorial was built on the Qualtrics survey platform. The tutorial was divided into single-screen units as to not overwhelm students (see recommendations for tutorial design; He *et al.*, 2012; Korkmaz and Harwood, 2004). Students were not allowed to return to previous units, as the pre- and post- tutorial assessment prompts were similar, and we did not want students to look back at their pre-submission prompts. Multiple choice items were given in each tutorial unit to engage respondents with the material and increase motivation to complete the tutorial per recommendation by successful tutorial developers (Allen, 1995; Dewald *et al.*, 2000; Dewald, 1999; Donaldson, 2000; Lo and Dale, 2009) .

Immediate feedback was used in the form of try-again feedback. In our tutorial, students were forced to choose the correct answers to all questions in order to proceed to the next tutorial unit. This procedure relies on the benefits of correct response feedback, while also requiring students to return to the question and choose the correct answer (Clariana *et al.*, 1991). While it has been found that elaborated feedback can be

المنارة للاستشارات
www.manaraa.com

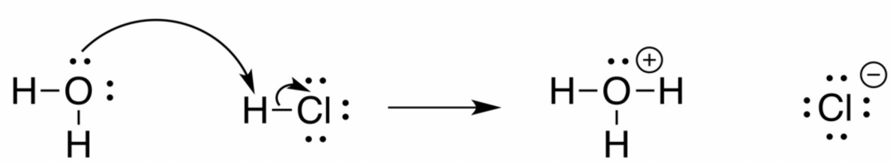
more effective than the other types of feedback (Jaehnig and Miller, 2007), the medium used (i.e., Qualtrics) did not allow for such elaborated feedback.

Each tutorial unit was designed to cover a different aspect or dimension of the Lewis acid–base model found in the literature to be difficult for postsecondary students. The tutorial, as a whole, was expected to increase the number of responses to our acid–base proton-transfer prompt that discuss the reaction using the Lewis acid–base model with an emphasis on the importance of electrons in the forming and breaking of bonds. Concepts associated with the Arrhenius and Brønsted-Lowry acid–base models were not specifically addressed in the tutorial; as such, we expected use of these models to be unaffected or decrease in post-tutorial responses to our prompt.

Pre-Tutorial Assessment Prompt.

Before the first tutorial unit, respondents were presented with an acid–base proton-transfer reaction between water and hydrochloric acid. This item was used in our prior work (Dood *et al.*, 2018) and adapted from an item first developed by Cooper *et al.* (2016). Respondents were asked to write about what is happening in the reaction (Part A) and why the reaction is happening (Part B; see Figure 4.1).

Consider the mechanism below for the acid-base reaction between water and hydrochloric acid to form hydronium ion and chloride ion.



A. Describe in full detail *what* you think is happening on the molecular level for this reaction. Be sure to discuss the role of each reactant.

B. Using a molecular level explanation, please explain *why* this reaction occurs. Be sure to discuss why the reactants form the products shown.

Figure 4.1. Pre-tutorial assessment prompt

المنارة للاستشارات
www.manaraa.com

Tutorial Unit 1: Definition of Lewis Acids and Bases.

The first tutorial unit (see Figure 4.2) introduced the definition of Lewis acids and bases in the context of the simple acid−base proton-transfer reaction written about in the pre-tutorial assessment prompt. Cartrette and Mayo (2011) found that more than half of the students they interviewed could not correctly define Lewis acids and bases nor apply the model. Keeping Cartrette and Mayo's results in mind, this tutorial unit sought to instill a clear definition of the Lewis model. The specific learning outcomes for this unit were to (A) define a Lewis acid and a Lewis base, and (B) identify which substance in an acid−base reaction is the Lewis acid and which substance is the Lewis base. These outcomes were assessed by asking respondents to identify the Lewis acid and the Lewis base in the reaction (see Figure 4.3). The first item was a simple proton transfer reaction, while the second item was more complex, containing a diprotic acid.

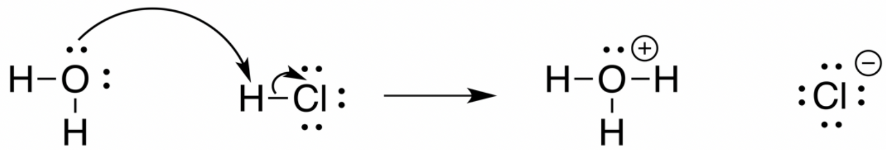
Tutorial Unit 2: Amphoteric Property of Water.

In the second tutorial unit, the amphoteric property of water was introduced. Note that water was a reactant in the Pre-Tutorial Assessment Prompt. The tutorial states: "Some chemical species can act as either a Lewis acid or a Lewis base; thus, these chemical species can either accept a pair of electrons or donate a pair of electrons. Water is an example of one such chemical species."

This definition is followed by three assessment items in which students must select whether water is acting as a Lewis acid or a Lewis base (see Figure 4.4). This unit addressed the difficulties identified by Garnett *et al.* (1995) and Drechsler and Van Driel (2008) regarding identifying the role that amphoteric species such as water play in acid−base reactions. The learning outcome for this tutorial unit was to identify whether water is acting as a Lewis acid or a Lewis base in a given reaction. This outcome was assessed by asking students to identify if water is acting as a Lewis acid or a Lewis base in three reactions (see Figure 4.4).

المنارة للاستشارات
www.manaraa.com

On the previous screen, you were asked to respond to a writing prompt that asked you to describe what was happening and why this reaction occurred.



You could have responded to the question using one of several acid-base models. Example models: Arrhenius, Brønsted-Lowry, or Lewis.

A key acid-base model is the Lewis acid-base model.
The Lewis model is focused on the transfer of electrons between molecules.

<u>Key Definitions</u>

- **A Lewis acid is an electron pair acceptor.**
- **A Lewis base in an electron pair donor.**

For example, in the reaction above, hydrochloric acid (HCl) would be the Lewis acid because it accepts the electron pair from water (H_2O), which leads to the bond between hydrogen and chloride breaking. Water would be considered a Lewis base because it donates an electron pair to form the bond between oxygen and hydrogen.

Figure 4.2. "Definition of Lewis Acids and Bases" tutorial unit

Tutorial Unit 3: Acid–Base Reactions Without a Proton Transfer.

The third tutorial unit introduced Lewis acid and base reactions that do not involve a proton-transfer, thus introducing contexts in which the Brønsted-Lowry model is inappropriate. The text of this unit is shown in Figure 4.5. The instruction and assessments given in this unit attempted to remedy the misconception of electron sharing in acid–base reactions as identified by Tarhan and Sesen (2012). The learning outcome for this unit was for students to identify Lewis acids and bases in a reaction that does not involve the transfer of a proton. The outcome was assessed by asking respondents to identify the Lewis acid and the Lewis base in three reactions that do not involve the transfer of a proton (see Figure 4.6 for the reactions).

www.manaraa.com

Based on the definitions of a Lewis acid and a Lewis base, consider this new acid-base reaction:

$$NH_3 + H_3O^+ \rightarrow NH_4^+ + H_2O$$

Which species is the Lewis acid?
○ NH_3
○ H_3O^+

Which species is the Lewis base?
○ NH_3
○ H_3O^+

Consider another acid-base reaction:

$$CH_3CH_2OH_2^+ + HSO_4^- \rightarrow H_2SO_4 + CH_3CH_2OH$$

Which species is the Lewis acid?
○ $CH_3CH_2OH_2^+$
○ HSO_4^-

Which species is the Lewis base?
○ $CH_3CH_2OH_2^+$
○ HSO_4^-

Figure 4.3. Example assessments for "Definition of Lewis Acids and Bases" tutorial unit

$$H_2O + NH_3 \rightarrow OH^- + NH_4^+$$

$$H_2O + HCl \rightarrow H_3O^+ + Cl^-$$

$$H_2O + HNO_3 \rightarrow NO_3^- + H_3O^+$$

Figure 4.4. Reactions for assessment of "Amphoteric Property of Water" tutorial unit

All of the acid-base reactions that have appeared in this tutorial so far have involved the transfer of a proton. However, Lewis acid/base reactions do not always include the transfer of protons.

Lewis acid/base reactions, in general, involve the formation of bonds involving an electron sufficient species (i.e., a Lewis base) and an electron deficient species (i.e., a Lewis acid).

Figure 4.5. Text of "Acid–Base Reactions Without a Proton Transfer" tutorial unit

Tutorial Unit 4: Integrating Concepts.

The final tutorial unit was focused on integrating the concepts covered: an item without a proton-transfer, an item with an amphoteric molecule, and two items involving a proton transfer. Tasks that require integrating

المنارة للاستشارات
www.manaraa.com

$$BF_3 + F^- \rightarrow BF_4^-$$

$$H_3N + BF_3 \rightarrow H_3N\text{–}BF_3$$

$$AlCl_3 + NH_3 \rightarrow Cl_3Al\text{–}NH_3$$

Figure 4.6. Reactions for assessment of “Acid−Base Reactions Without a Proton Transfer” tutorial unit

multiple concepts allows students to construct more expert-like understanding (Taagepera and Noori, 2000). The unit text states, “Based on the definition of Lewis acid−base reactions that we have reviewed in this tutorial, determine what species is the Lewis acid AND what species is the Lewis base for these reactions.” The learning outcome was for students to apply the knowledge they have constructed via the tutorial to a broader spectrum of chemical contexts (i.e., charged species, multiple acids, and multi-protic acids). This learning outcome was assessed by asking respondents to identify the Lewis acid and the Lewis base in a series of reactions (see Figure 4.7 for reactions).

$$(CH_3)_3C^+ + Br^- \rightarrow (CH_3)_3CBr$$

$$CH_3COOH + H_2O \rightarrow CH_3COO^- + H_3O^+$$

$$H_3PO_4 + CH_3OH \rightarrow H_2PO_4^- + CH_3OH_2^+$$

Figure 4.7. Reactions for the assessment of “Integrating Concepts” tutorial unit

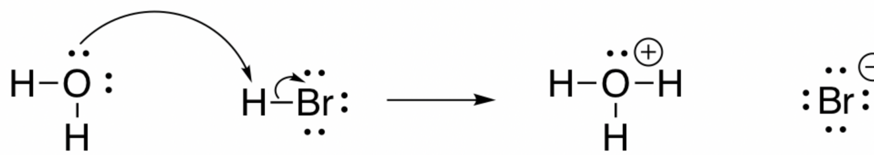
Post-Tutorial Assessment Prompt.

Finally, students responded to a second writing prompt (see Figure 4.8). This prompt was a “cloned” version of the first prompt, i.e., the two questions were similar, the acid and conjugate base were changed. Cloned items have been used in learning environments such that students can retake multiple cloned items until a

المنارة للاستشارات
www.manaraa.com

desired learning outcome is mastered (Lathrop and Cheng, 2017). Cloned items can be used in pre-test/post-test experimental designs to evaluate interventions (Lathrop and Cheng, 2017). The pre-tutorial assessment prompt included hydrochloric acid, while the post-tutorial assessment prompt included hydrobromic acid.

Consider the mechanism below for the acid-base reaction between water and hydrobromic acid to form hydronium ion and bromide ion.



C. Describe in full detail *what* you think is happening on the molecular level for this reaction. Be sure to discuss the role of each reactant.

D. Using a molecular level explanation, please explain *why* this reaction occurs. Be sure to discuss why the reactants form the products shown.

Figure 4.8. Post-tutorial assessment prompt

4.9.2 Expansion of the predictive logistic regression model

We previously reported a predictive logistic regression model for determining Lewis acid−base model usage in student responses to an acid−base proton-transfer (Dood *et al.*, 2018). However, this model is limited to predicting Lewis model use for the reaction of hydrochloric acid with water. To evaluate the impact of the tutorial, we needed to predict Lewis model use with the original reaction and a "cloned" version of the reaction (i.e., hydrobromic acid and water). In addition, we evaluated the reaction of hydroiodic acid and water to demonstrate the broad utility of the logistic regression model for predicting a broader range of hydrohalic acids and water reactions.

Data were collected in the first semester of a two-semester organic chemistry course during the spring 2018 semester at a large, public university in the southeastern United States. The assessment was given as an extra credit opportunity. The survey, given via Qualtrics, was programmed such that half of the students

المنارة للاستشارات
www.manaraa.com

would receive the version of the question with hydrobromic acid and the other half would receive hydroiodic acid. A total of 306 usable responses were received, with 152 students receiving the version of the question with hydrobromic acid, and 154 students receiving the version of the question with hydroiodic acid. These responses were hand-coded and computer-coded for the use of a Lewis acid−base model in the response.

Coding methods for the original predictive model (Dood *et al.*, 2018) were edited to account for the new acids (i.e., HI and HBr) and conjugate bases (i.e., I^- and Br^-). As an example, every instance where the type "chlorine" was used in the model was replaced by a new type called "conjugate base" which included both "iodine" and "bromine", along with common incorrect spellings. (See Dood *et al*. for a complete discussion of the coding method.)

The expanded logistic regression model predicted Lewis acid−base usage in student responses with 86% accuracy (i.e., greater than the original accuracy of 82%), indicating that the model can predict Lewis acid−base model usage for reactions with different hydrohalic acids. This broader model was then used when testing the tutorial model.

4.9.3 Testing the effectiveness of the tutorial

The effectiveness of the tutorial was tested in three instructional settings within a yearlong second-year organic chemistry sequence. The tutorial was tested with students in their first semester before summative assessment on acids and bases (n = 457, Spring 2019), students in their first semester following summative assessment on acids and bases (n = 348, Fall 2018), and students in their second semester (n = 210, Fall 2018). Additionally, the students in their first semester of organic chemistry in Spring 2019 were reassessed using the prompt three weeks after the initial assessment and tutorial (n = 427). Testing was done with students who had three different instructors at a large, public university in the southeastern United States. In all cases, the tutorial assignment was given as an extra credit opportunity. In the tutorial, students answered the proton-transfer acid−base prompt using hydrochloric acid. Then, students were given

www.manaraa.com

the Lewis acid–base tutorial. Finally, students answered a cloned version of the proton-transfer acid–base prompt with hydrobromic acid. The predictive logistic regression model was used to assign Lewis model "use" (score = *1*) or "non-use" (score = *0*) for the pre- and post-tutorial constructed-responses. A McNemar test for comparison of correlated proportions (McNemar, 1947) was conducted to compare "use" or "non-use" of the Lewis acid–base model in pre- and post-tutorial written responses (α = 0.05) for each instructional setting. A McNemar test was also conducted to compare "use" or "non-use" of the Lewis acid–base model in post-tutorial written responses and written responses after a three-week time delay. Additionally, categories used in the predictive logistic regression model were evaluated pre/post tutorial using the McNemar test for the combined data set including all three instructional contexts; α was adjusted to 0.0027 to account for multiple comparisons (i.e., Type 1 error).

The output of a McNemar test includes χ^2, p, and an odds ratio. The McNemar test considers two groups with different outcomes, one that is an ideal outcome and one that is a non-ideal outcome. In this study, the ideal group is the students who are coded as "use" of the Lewis acid–base model and the non-ideal group is the students who are coded as "non-use" of the Lewis acid–base model. Both groups receive the treatment (i.e., the tutorial). After the treatment, the two groups are determined and compared to the groups before the treatment occurred. The odds ratio (Cleophas and Zwinderman, 2016; Sheskin, 2011) is calculated by taking the number of subjects who moved from the non-ideal outcome to the ideal outcome (i.e., from "non-use" to "use" of the Lewis acid–base model) divided by the number of subjects who moved from the ideal outcome to the non-ideal outcome (i.e., from "use" to "non-use" of the Lewis acid–base model).

المنارة للاستشارات
www.manaraa.com

4.10 Results

The goal of our study was to evaluate the effectiveness of a tutorial designed to increase student usage of the Lewis acid−base model in their responses to what is happening and why in an acid−base proton-transfer reaction. Students responded to a writing prompt before and after completing the Lewis acid−base tutorial. These responses were evaluated to determine the impact of the tutorial.

4.10.1 First semester students before summative assessment

The results of a McNemar test for the responses of first semester students before summative assessment indicate a significant increase in usage of the Lewis acid−base model in student responses post-tutorial (χ^2 = 20.2, p = <0.00001, OR = 2.3). Before the tutorial, 69.8% of student responses were coded as "use" of the Lewis acid−base model. After the tutorial, 80.9% of student responses were coded as "use" of the Lewis acid−base model, a net increase of 11.1% post-tutorial. Of the students who were coded as "non-use" pre-tutorial, 65.2% were coded as "use" after receiving the tutorial. Table 4.1 shows a cross tabulation of responses coded as "use" and "non-use" of the Lewis acid−base model before and after being exposed to the tutorial.

Table 4.1 Cross-tabulation of responses coded as "use" and "non-use" pre- and post-tutorial for first-semester students before summative assessment

	Pre-tutorial		
Post-tutorial	"Non-use"	"Use"	Total
"Non-use"	48	39	87
"Use"	90	280	370
Total	138	319	457

المنارة للاستشارات
www.manaraa.com

4.10.2 Post-tutorial assessment after time delay

The results of a McNemar test for the responses of the first semester students pre-summative assessment and post-tutorial compared to the responses of the same students post-tutorial after a three-week time delay indicate a non-significant decrease in use of the Lewis acid−base model (χ^2 = 3.33, p = 0.0679, OR = 0.71). Immediately following the tutorial, 81.5% of student responses were coded as "use" of the Lewis acid−base model; this percentage differs from the post-tutorial percent given above because not all students completed the post-tutorial assessment after the time delay. After a three-week delay, 76.8% of student responses were coded as "use" of the Lewis acid−base model, a decrease of 4.7% following the three-week time delay. Table 4.2 shows a cross tabulation of responses coded as "use" and "non-use" of the Lewis acid−base model before and after being exposed to the tutorial.

Table 4.2 Cross-tabulation of responses coded as "use" and "non-use" immediately post-tutorial and after a 3-week time delay for first-semester students before summative assessment

	Post-tutorial		
After time delay	"Non-use"	"Use"	Total
"Non-use"	29	70	99
"Use"	50	278	328
Total	79	348	427

4.10.3 First semester students after summative assessment

The results of a McNemar test for the responses of first semester students after summative assessment also indicate an increase in usage of the Lewis acid−base model in student responses post-tutorial (χ^2 = 15.1, p = 0.0001, OR = 2.52). Before the tutorial, 73.2% of student responses were coded as "use" of the Lewis acid−base model. After the tutorial, 83.3% of students were coded as "use" of the Lewis acid−base model, a net increase of 10.1%. Of the students who were coded as "non-use" pre-tutorial, 62.3% were

المنارة للاستشارات
www.manaraa.com

coded as "use" after receiving the tutorial. Table 4.3 shows a cross tabulation of responses coded as "use" and "non-use" of the Lewis acid−base model before and after being exposed to the tutorial.

Table 4.3 Cross-tabulation of responses coded as "use" and "non-use" pre- and post-tutorial for first-semester students after summative assessment

	Pre-tutorial		
Post-tutorial	"Non-use"	"Use"	Total
"Non-use"	35	23	58
"Use"	58	232	290
Total	93	255	348

4.10.4 Second semester students

The results of a McNemar test for the responses of second semester students indicate a significant increase in usage of the Lewis acid−base model post-tutorial ($\chi^2 = 9.29$, $p = 0.0023$, OR = 2.26). Before the tutorial, 71.9% of student responses were coded as "use" of the Lewis acid−base model. After the tutorial, 83.3% of students were coded as "use" of the Lewis acid−base model, a net increase of 11.4% post-tutorial. Of the students who were coded as "non-use" pre-tutorial, 72.9% were coded as "use" post-tutorial. Table 4.4 shows a cross tabulation of responses coded as "use" and "non-use" of the Lewis acid−base model before and after being exposed to the tutorial.

Table 4.4 Cross-tabulation of responses coded as "use" and "non-use" pre- and post-tutorial for second-semester students

	Pre-tutorial		
Post-tutorial	"Non-use"	"Use"	Total
"Non-use"	16	19	35
"Use"	43	132	175
Total	59	151	210

المنارة للاستشارات
www.manaraa.com

4.11 Discussion

The tutorial showed similar impact in all three instructional settings, indicating that the tutorial has utility in different instructional contexts. Of note, 81 total students from all three contexts combined moved from "use" to "non-use" following the tutorial (n = 725); this number is substantially smaller than number of students who were coded as "non-use" before the tutorial and moved to "use" after the tutorial (n= 191 of 290). Additionally, a closer look at the responses of these students who appear to have regressed revealed responses that included Lewis acid−base related concepts, such as the transfer of electrons, but also included other factors that were strong negative predictors, such as "dissociation" and "strong acid", which caused the response to be coded overall as "non-use". Some regression can also be accounted for by the error in the predictive model (i.e., either the first prompt returned a false positive or the second prompt returned a false negative). Students experiencing fatigue (Lavrakas, 2008) by the end of the assessment may also have contributed to this number, as the average time to complete the prompt was approximately 1.5 minutes shorter for the post-prompt than for the pre-prompt. While these regressions are not ideal, a closer look revealed sufficient evidence that factors beyond Lewis acid−base model use influenced students' responses.

A decay of knowledge regarding "use" of the Lewis acid−base was observed when students were given the assessment prompt after a three-week time delay. However, the change in responses between immediately post-tutorial and after a three-week time delay was not significant according the McNemar test. Some decay in knowledge is expected, as chemistry students have been found to experience a significant decrease in achievement following summative assessment (Bunce *et al.*, 2011); this is evidenced by the decrease in "use" of the Lewis acid−base model by 4.7%. This indicates a decrease in "use" of the Lewis acid−base model in the three weeks since the tutorial, but a net increase between the pre-tutorial prompt and the prompt given following a three-week time delay. Another possible explanation for the sudden "post" increase with an additional decrease "post-post" is that the students memorized the desired explanation and

المنارة للاستشارات
www.manaraa.com

regurgitated it in the "post" assessment without meaningful integration of the concepts; while we cannot rule out this approach, the net increase after a three-week time delay suggests that some students did approach the tutorial meaningfully. We cannot claim that the lasting impact was due to the tutorial alone; students had experienced three additional weeks of instruction in organic chemistry, which could have also been a contributing factor to their "use" of the Lewis acid−base model. Also, students were not assessed on acids and bases during the delay time; it is therefore possible that there was some lasting impact of the tutorial. The tutorial was found to have a net positive impact as a review tool for students in their first semester of organic chemistry after summative assessment and for students in their second semester of organic chemistry. The utility of the tutorial is broader than for use during acid−base instruction, as the tutorial can be repeated and maintain a positive impact on Lewis acid−base model use.

The tutorial was designed to address the definitions of Lewis acids and bases (i.e. electron acceptors and donors) and promote use of the Lewis acid−base model. Categories from the predictive model were analyzed to determine if the number of students coded in each category was significantly different after the tutorial. Specifically, we are interested in categories associated with the Lewis acid−base model. However, we additionally explored whether categories associated with the other acid−base models (i.e., Brønsted-Lowry and Arrhenius) stayed the same or decreased in coded frequency. Results of McNemar tests for all categories are given in Table 4.2.

Categories related to the use of a Lewis acid−base model show significant increases following the tutorial. These include *nucleophile or electrophile*, *donate electrons*, *accept electrons*, and *oxygen as an electron donor*. Looking at the odds ratios, the odds of using these categories were 17.1, 2.6, 2.4, and 1.7 times higher following the tutorial when holding all other factors constant, respectively. We expected the categories related to the transfer of electrons to be used by students after receiving the tutorial due to this being the focus of the tutorial. The category *electron pairs* also showed a positive odds ratio, but its change

www.manaraa.com

Table 4.5 Results of the McNemar's test for predictive logistic regression model categories

Category (model)	χ^2	p	Odds^a	Pre "use" (%)	Post "use" (%)
Nucleophile or electrophile (L)	387	<0.001*	17.1	11.5	54.4
Donate electrons (L)	74.9	<0.001*	2.64	38.4	54.8
Accept electrons (L)	68.8	<0.001*	2.36	35.2	51.9
Oxygen as an electron donor (L)	19.6	<0.001*	1.68	20.4	28.0
Electron pairs (L)	0.43	0.510	1.09	76.3	77.2
Donate protons (B)	2.06	0.151	1.16^a	41.5	38.7
Bond electrons (L)	2.49	0.115	1.21^a	25.1	22.6
Conjugate acid or base (B)	6.95	0.008	1.46^a	23.6	20.0
Forming ions (A)	10.8	0.001*	1.61^a	29.2	24.7
Hydrogen and not electrons (B)	9.98	0.002*	1.72^a	20.7	16.4
Partial charges (L)	8.67	0.003	2.21^a	7.0	4.7
Hydrogen actions (B)	52.2	<0.001*	2.36^a	52.7	40.0
Accept protons (B)	59.5	<0.001*	2.52^a	77.8	64.2
Sharing electrons (L)	20.3	<0.001*	2.79^a	10.2	6.0
Acid strength (A)	28.5	<0.001*	3.07^a	13.6	8.1
Electronegativity (L)	38.5	<0.001*	3.62^a	14.5	7.9
Attraction of hydrogen (B)	59.3	<0.001*	4.06^a	20.7	11.1
Dissociation (A)	43.6	<0.001*	4.88^a	14.4	7.9

Acid−base models: L, Lewis; B, Brønsted-Lowry; A, Arrhenius. *$p < 0.0027$. ^{a}inverse odds ratio.

in use was not significant at the 0.0027 level. The increase in use of the category *nucleophile or electrophile* appears to not fit given the content of the tutorial, but the category includes use of the terms "Lewis base" and "Lewis acid" (Dood *et al.*, 2018) as the terms are synonymous with nucleophiles and electrophiles in this context. Considering a negative impact of the tutorial, the categories "electronegativity" and "sharing electrons" were 3.6 and 2.8 times less likely to be used following the tutorial when holding all other factors constant; though these concepts are associated with a Lewis acid−base model, we hypothesize that the use of both categories decreased following the tutorial due to the concepts not being explicitly addressed in the tutorial.

"Non-use" of a Lewis acid−base model may indicate to some degree the "use" of a Brønsted-Lowry model or an Arrhenius model in the student's explanation, as the Brønsted-Lowry and Arrhenius models are alternative ways to describe what is happening in the acid−base proton-transfer reaction and why; mixed-model use was found in our original study (Dood *et al.*, 2018). Many categories related to the Brønsted-Lowry and Arrhenius models showed a significant decrease in use post-tutorial. Of the categories with

المنارة للاستشارات
www.manaraa.com

a significant difference before and after the tutorial, *accept protons*, *hydrogen actions*, *hydrogen and not electrons*, and *attraction of hydrogen* are related to use of the Brønsted-Lowry model. For example, a student would likely discuss the reaction in terms of hydrogen and not include electrons in their discussion when invoking solely a Brønsted-Lowry model. The odds of using categories *accept protons*, *hydrogen actions*, *hydrogen and not electrons*, and *attraction of hydrogen* were 2.5, 2.4, 1.7, and 4.1 less lower for responses after the tutorial, respectively, when holding all other factors constant. The categories *donate proton* and *conjugate acid or base* also relate to Brønsted-Lowry acids and bases (i.e., a Brønsted acid is a proton donor, a Brønsted base is a proton acceptor, a reaction of a Brønsted-Lowry acid, and a Brønsted-Lowry base results in a conjugate acid and a conjugate base) and showed a negative odds ratios, but they were not significant at the 0.0027 level. We did not expect a significant change in use of constructs related to the Brønsted-Lowry model as the tutorial did not encourage respondents not to use the model. However, it appears that promotion of the Lewis acid−base model in the tutorial, in turn, suppressed use of the Brønsted-Lowry model. Categories *forming ions*, *dissociation*, and *acid strength* are strongly associated with the Arrhenius acid−base model, and had 1.6, 4.9, and 3.1 times lower odds of use used following the tutorial when holding all other factors constant, respectively.

The tutorial moved many students from “non-use” of a Lewis model to “use” of a Lewis model when describing what happens in an acid−base proton-transfer reaction and why. Some students (11.6% of all students given the tutorial) regressed from the “use” of a Lewis acid−base model to “non-use” of a Lewis acid−base model. This is less than the number of students who moved from “non-use” to “use” (18.1%), which implies that the tutorial is an overall effective learning tool for helping students construct a meaningful understanding of the use and application of the Lewis model in the contexts we provided.

The high number of students coded as “use” of a Lewis acid−base model pre-tutorial creates the possibility of the occurrence of a ceiling effect. Though only 18.1% of students across instructional contexts

المنارة للاستشارات
www.manaraa.com

moved from “non-use” to “use” of the Lewis acid−base model, if the number of students not using the Lewis acid−base model pre-tutorial had been lower, a greater impact of the tutorial may have been observed.

The number of students coded as “use” of the Lewis acid−base model pre-tutorial was similar across contexts (i.e., 69.8% for first semester pre-summative assessment, 73.3% for first semester post-summative assessment, and 71.9% for second semester). This indicates that student usage of the model may not be increasing as they progress through an organic chemistry course. This provides further evidence of the need for interventions that can be used repeatedly throughout the duration of a yearlong course in organic chemistry to serve as “refreshers” for students to promote lasting understanding of topics covered at the beginning of the course.

4.12 Implications for research and teaching

Using this study and associated predictive model as a framework, more tutorial tools should be developed to address known issues with student understanding of reactions and reaction mechanisms. We have shown that the predictive model for the acid−base proton-transfer reaction can be used to evaluate student responses in a pre/post-tutorial context. Further predictive models should be developed for scoring constructed-response items that focus on a broader range of reactions (e.g., S_N1, S_N2, E1, E2). These models should then be used to evaluate tutorials that address aspects of the other reaction types known to be difficult. Such work should be built using the vast array of literature that exists regarding student understanding of organic chemistry reaction mechanisms (Bhattacharyya, 2014; Bhattacharyya and Bodner, 2005; Bhattacharyya and Harris, 2018; Ferguson and Bodner, 2008; Flynn and Featherstone, 2017; Flynn and Ogilvie, 2015; Galloway *et al.*, 2017; Grove *et al.*, 2012a, 2012b). For example, research findings related to student understanding of other reaction types and components of reactions (e.g., nucleophiles and

المنارة للاستشارات
www.manaraa.com

electrophiles; Anzovino and Bretz, 2015, 2016; leaving groups; Popova and Bretz, 2018) could be used to develop additional tutorials.

Instructors can use this tutorial in their classes as a way to teach the Lewis acid–base model or as a tool for students to solidify their understanding following instruction. As we have shown, the tutorial can have a net positive impact on students in their second semester of organic chemistry. Thus, the tutorial could also be used as a way for students to review Lewis acid–base concepts at the beginning of their second semester of organic chemistry. In addition to effectively helping students build an understanding of the Lewis acid–base model, instructors can receive feedback from the writing portion of the tutorial about whether or not their students are invoking the Lewis model when describing the acid–base proton transfer reaction. Instructors can also explore the specific categories that their students are using in their responses and potentially address areas where students have weak understanding during lectures. Future versions of the tutorial could include immediate feedback for students regarding the use of the Lewis acid–base model and other categories in their writing following the second writing prompt; this would allow students to identify ideas that they are missing in order to develop a more complete explanation of acid–base proton-transfer reactions.

4.13 Limitations

The main limitation of our study is that it lacks a control group, i.e. all students were given the tutorial. There is an ethical concern regarding giving one group of students a tutorial that we expect to be helpful to their learning and withholding the tutorial from another group of students. The use of a "distractor task" given between the two iterations of the cloned prompt is one method for controlling for the impact of an intervention; however such a method is more appropriate in clinical research settings (Armitage *et al.*, 2014; Butler *et al.*, 2007; Wang and Thomas, 1995). Thus, our choice of statistical methods (i.e., the

المنارة للاستشارات
www.manaraa.com

McNemar test) is purposeful because the technique can be used to evaluate cause and effect impact when having a control group is unethical or impractical for evaluating an intervention (McNEmar, 1947; Sheskin, 2011).

Additionally, the current platform we use to administer the prompts and tutorial, i.e. Qualtrics, does not allow for targeted feedback when specific incorrect answers are selected. In the future, the tutorial could be moved to a platform which allows for targeted feedback by answer option. We expect that such feedback would further the impact of the tutorial.

Another limitation is the need for computer scoring of the written response outside of the tutorial. Development of an API (application program interface) to return the computer scored item would allow for targeted tutorial units to be shown or not shown to the respondent based on computer-scored proficiency. For example, if a student sufficiently demonstrated an understanding of “accept” and “donate” electrons in their pre-tutorial assessment, the tutorial platform may truncate several units to quickly reinforce the Lewis acid−base model and end the tutorial.

Despite the imperfect accuracy of the predictive model, changes were observed in student responses after the tutorial. The 86% accuracy, however, may limit our measure of ‘true’ effectiveness. While the tool is sufficient for formative assessment where mischaracterization by the computer is inconsequential to the student, we recommend against using the tool or tutorial for summative assessment purposes where a student’s performance in the course is dependent on the score returned by the predictive model.

No long-term gains were measured following the tutorial. We maintain that the tutorial is still useful, as we saw immediate gains in all contexts and the tutorial could be used as a learning tool and a refresher tool. We also saw immediate gains for students using the tutorial post-assessment in their first semester of organic chemistry and in their second semester of organic chemistry, indicating that the tutorial has a positive impact for “use” of the Lewis acid−base model if repeated at different time points. Repetition of concepts

www.manaraa.com

and multiple assessment opportunities have been shown to remediate the decay in student knowledge that occurs in chemistry (Bunce *et al.*, 2011).

The assessment prompts include monoprotic halogenic acids. In the future, other acids should be added to our model to broaden the scope, including di- and triprotic acids, such as H_2SO_4 and H_3PO_4. The scope should also be broadened by adding other bases, for example alcohols or amines. In order to provide students with a more complete understanding of the Lewis acid−base model, prompts should be developed for non-protic acids and a wider array of bases.

The current model predicts "use" or "non-use" of the Lewis acid−base model and does not evaluate correctness of such use. In our previous work, students whose responses were coded as "use" according to the same coding scheme scored statistically higher on a summative organic chemistry assessment that was scored for correctness (Dood *et al.*, 2018). This indicates that use of Lewis acid−base model terminology can be helpful for students, regardless of the correctness of use.

The tutorial does not address the use of causal reasoning to describe why the acid−base reaction is occurring. Work by Cooper *et al.* (2016) suggests that students who engage in both the use of a Lewis acid−base model and causal reasoning have the greatest success in organic chemistry. However, the use of the Lewis acid−base model in the absence of causal reasoning has been shown to be associated with success in producing the mechanism for the acid−base proton-transfer reaction, indicating that simply promoting Lewis acid−base model use can have an impact on student success in organic chemistry; in our previous work (Dood *et al.*, 2018), we did not evaluate the use of causal mechanistic reasoning, but saw that students who used the Lewis acid−base model were more successful on an examination, indicating again that a tutorial promoting the use of a Lewis acid−base model can be helpful for students even if it does not also address causal reasoning. A future tutorial to be used by students after this tutorial could be used to promote causal mechanistic reasoning in relation to the acid−base proton-transfer reaction presented in this manuscript.

المنارة للاستشارات
www.manaraa.com

4.14 Conclusion

A firm understanding of the Lewis acid–base model is important for learning many concepts in organic chemistry. Students struggle to apply the Lewis acid–base model, having more experience and confidence with the Brønsted-Lowry model (Cartrette and Mayo, 2011). Tutorial activities have shown to be helpful in learning chemistry. We have presented and determined the effectiveness of a tutorial related to Lewis acid–base theory that increased the number of students invoking a Lewis acid–base model in their post-tutorial assessments. In the future, the constructed-response item, coupled with resultant predictive models, should be incorporated into tutorials used for promoting learning of other organic chemistry reactions (e.g., S_N1, E1, EAS, etc.).

4.15 Acknowledgments

We thank those students who completed the tutorial as part of their organic chemistry coursework. Additionally, we gratefully acknowledge John Ferron (University of South Florida, Tampa, Florida, USA) for advice on the research design and statistical analyses used. Finally, we acknowledge John C. Dood (Tampa, Florida, USA) for his expertise with Python and troubleshooting our predictive model program.

4.16 References

Allen, E.E., (1995), Active Learning and Teaching: *Ref. Libr.*, **24**, 89–103.

Anzovino, M.E., Bretz, S.L., (2016), Organic chemistry students' fragmented ideas about the structure and function of nucleophiles and electrophiles: a concept map analysis. *Chem. Educ. Res. Pract.*, **17**, 1019–1029.

Anzovino, M.E., Bretz, S.L., (2015), Organic chemistry students' ideas about nucleophiles and electrophiles: the role of charges and mechanisms. *Chem. Educ. Res. Pract.*, **16**, 797–810.

المنارة للاستشارات
www.manaraa.com

Arasasingham, R.D., Martorell, I., McIntire, T.M., (2011), Online Homework and Student Achievement in a Large Enrollment Introductory Science Course. *J. Coll. Sci. Teach.*, **40**, 70–79.

Arasasingham, R.D., Taagepera, M., Potter, F., Martorell, I., Lonjers, S., (2005), Assessing the Effect of Web-Based Learning Tools on Student Understanding of Stoichiometry Using Knowledge Space Theory. *J. Chem. Educ.*, **82**, 1251–1262.

Armitage, C.J., Rowe, R., Arden, M.A., Harris, P.R., (2014), A brief psychological intervention that reduces adolescent alcohol consumption. *J. Consult. Clin. Psychol.*, **82**, 546–550.

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

Bhattacharyya, G., (2013), From source to sink: mechanistic reasoning using the electron-pushing formalism. *J. Chem. Educ.*, **90**, 1282–1289.

Bhattacharyya, G., Bodner, G.M., (2005), "It gets me to the product": how students propose organic mechanisms. *J. Chem. Educ.*, **82**, 1402–1407.

Bhattacharyya, G., Harris, M.S., (2018), Compromised structures: Verbal descriptions of mechanism diagrams. *J. Chem. Educ.*, **95**, 366–375.

Bretz, S.L., McClary, L., (2015), Students' understandings of acid strength: How meaningful is reliability when measuring alternative conceptions? *J. Chem. Educ.*, **92**, 212–219.

Bunce, D.M., VandenPlas, J.R., Soulis, C., (2011), Decay of student knowledge in chemistry. *J. Chem. Educ.*, **88**, 1231–1237.

Butler, A.C., Karpicke, J.D., Roediger III, H.L., (2007), The effect of type and timing of feedback on learning from multiple-choice tests. *J. Exp. Psychol. Appl.*, **13**, 273–281.

Cartrette, D.P., Mayo, P.M., (2011), Students' understanding of acids/bases in organic chemistry contexts. *Chem. Educ. Res. Pract.*, **12**, 29–39.

Clariana, R.B., Ross, S.M., Morrison, G.R., (1991), The effects of different feedback strategies using computer-administered multiple-choice questions as instruction. *Educ. Technol. Res. Dev.*, **39**, 5–17.

Cleophas, T.J., Zwinderman, A.H., (2016), McNemar's Odds Ratios, in: Cleophas, T.J., Zwinderman, A.H. (Eds.), Clinical Data Analysis on a Pocket Calculator: Understanding the Scientific Methods of Statistical Reasoning and Hypothesis Testing. Springer International Publishing, Cham, pp. 275–278.

Cooper, M., (2015), Why ask why? *J. Chem. Educ.*, **92**, 1273–1279.

Cooper, M., Kouyoumdjian, H., Underwood, S., (2016), Investigating students' reasoning about acid–base reactions. *J. Chem. Educ.*, **93**, 1703–1712.

Demircioglu, G., Ayas, A., Demircioglu, H., (2005), Conceptual change achieved through a new teaching program on acids and bases. *Chem. Educ. Res. Pract.*, **6**, 36–51.

Dempsey, J.V., (1993), Interactive Instruction and Feedback. Educational Technology.

Dewald, N., Scholz-Crane, A., Booth, A., Levine, C., (2000), Information literacy at a distance: instructional design issues. *J. Acad. Librariansh.*, **26**, 33–44.

Dewald, N.H., (1999), Transporting good library instruction practices into the web environment: An analysis of online tutorials. *J. Acad. Librariansh.*, **25**, 26–31.

Donaldson, K.A., (2000), Library research success: Designing an online tutorial to teach information literacy skills to first-year students. *Internet High. Educ.*, **2**, 237–251.

Dood, A.J., Fields, K.B., Raker, J.R., (2018), Using lexical analysis to predict Lewis acid–base model use in responses to an acid–base proton-transfer reaction. *J. Chem. Educ.*, **95**, 1267–1275.

Drechsler, M., Schmidt, H.-J., (2005), Textbooks' and teachers' understanding of acid−base models used in chemistry teaching. *Chem. Educ. Res. Pract.*, **6**, 19–35.

Drechsler, M., Van Driel, J., (2008), Experienced teachers' pedagogical content knowledge of teaching acid–base chemistry. *Res. Sci. Educ.*, **38**, 611–631.

المنارة للاستشارات
www.manaraa.com

Erduran, S., (2003), Examining the mismatch between pupil and teacher knowledge in acid−base chemistry. *Sch. Sci. Rev.*, **84/308**, 81–87.

Ferguson, R., Bodner, G.M., (2008), Making sense of the arrow-pushing formalism among chemistry majors enrolled in organic chemistry. *Chem. Educ. Res. Pract.*, **9**, 102–113.

Flynn, A.B., Featherstone, R.B., (2017), Language of mechanisms: exam analysis reveals students' strengths, strategies, and errors when using the electron-pushing formalism (curved arrows) in new reactions. *Chem. Educ. Res. Pract.*, **18**, 64–77.

Flynn, A.B., Ogilvie, W.W., (2015), Mechanisms before reactions: A mechanistic approach to the organic chemistry curriculum based on patterns of electron Ffow. *J. Chem. Educ.*, **92**, 803–810.

Freasier, B., Collins, G., Newitt, P., (2003), A web-based interactive homework quiz and tutorial package to motivate undergraduate chemistry students and improve learning. *J. Chem. Educ.*, **80**, 1344–1347.

Galloway, K.R., Stoyanovich, C., Flynn, A.B., (2017), Students' interpretations of mechanistic language in organic chemistry before learning reactions. *Chem Educ Res Pr.*, **18**, 353–374.

Garnett, Patrick J., Garnett, Pamela J., Hackling, M.W., (1995), Students' alternative conceptions in chemistry: A review of research and implications for teaching and learning. *Stud. Sci. Educ.*, **25**, 69–96.

Grove, N., Cooper, M., Cox, E., (2012), Does mechanistic thinking improve student success in organic chemistry? *J. Chem. Educ.*, **89**, 850–853.

Grove, N.P., Cooper, M.M., Rush, K.M., (2012), Decorating with arrows: Toward the development of representational competence in organic chemistry. *J. Chem. Educ.*, **89**, 844–849.

Ha, M., Nehm, R.H., (2016), The impact of misspelled words on automated computer scoring: a case study of scientific explanations. *J. Sci. Educ. Technol.*, **25**, 358–374.

Ha, M., Nehm, R.H., Urban-Lurain, M., Merrill, J.E., (2011), Applying computerized-scoring models of written biological explanations across courses and colleges: prospects and limitations. *CBE-Life Sci. Educ.*, **10**, 379–393.

Haudek, K.C., Prevost, L.B., Moscarella, R.A., Merrill, J., Urban-Lurain, M., (2012), What are they thinking? Automated analysis of student writing about acid−base chemistry in introductory biology. *CBE Life Sci. Educ.*, **11**, 283–293.

He, Y., Swenson, S., Lents, N., (2012), Online video tutorials increase learning of difficult concepts in an undergraduate analytical chemistry course. *J. Chem. Educ.*, **89**, 1128–1132.

Herron, F.Y., (1953), Models illustrating the Lewis theory of acids and bases. *J. Chem. Educ.*, **30**, 199–201.

Jaehnig, W., Miller, M.L., (2007), Feedback types in programmed instruction: A systematic review. *Psychol. Rec.*, **57**, 219–232.

Korkmaz, A., Harwood, W.S., (2004), Web-supported chemistry education: Design of an online tutorial for learning molecular symmetry. *J. Sci. Educ. Technol.*, **13**, 243–253.

Kousathana, M., Demerouti, M., Tsaparlis, G., (2005), Instructional misconceptions in acid−base equilibria: an analysis from a history and philosophy of science perspective. *Sci. Educ.*, **14**, 173–193.

Kulik, J.A., Kulik, C.-L.C., (1988), Timing of feedback and verbal learning. *Rev. Educ. Res.*, **58**, 79–97.

Lathrop, Q.N., Cheng, Y., (2017), Item cloning variation and the impact on the parameters of response models. *Psychometrika*, **82**, 245–263.

Lin, J.-W., Chiu, M.-H., (2010), The mismatch between students' mental models of acids/bases and their sources and their teacher's anticipations thereof. *Int. J. Sci. Educ.*, **32**, 1617–1646.

Littlejohn, A., Suckling, C., Campbell, L., McNicol, D., (2002), The amazingly patient tutor: students' interactions with an online carbohydrate chemistry course. *Br. J. Educ. Technol.*, **33**, 313–321.

Lo, L.S., Dale, J.M., (2009), Information literacy "learning" via online tutorials: A collaboration between subject specialist and instructional design librarian. *J. Libr. Inf. Serv. Distance Learn.*, **3**, 148–158.

Luder, W.F., (1948), Contemporary acid−base theory. *J. Chem. Educ.*, **25**, 555–558.

McClary, L., Bretz, S.L., (2012), Development and assessment of a diagnostic tool to identify organic chemistry students' alternative conceptions related to acid strength. *Int. J. Sci. Educ.*, **34**, 2317–2341.

للاستشارات

www.manaraa.com

McClary, L., Talanquer, V., (2011a), College chemistry students' mental models of acids and acid strength. *J. Res. Sci. Teach.*, **48**, 396–413.

McClary, L., Talanquer, V., (2011b), Heuristic reasoning in chemistry: making decisions about acid strength. *Int. J. Sci. Educ.*, **33**, 1433–1454.

McNemar, Q., (1947), Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, **12**, 153–157.

Nakhleh, M.B., Krajcik, J.S., (1994), Influence of levels of information as presented by different technologies on students' understanding of acid, base, and pH concepts. *J. Res. Sci. Teach.*, **31**, 1077–1096.

Nataro, C., Ferguson, M.A., Bocage, K.M., Hess, B.J., Ross, V.J., Swarr, D.T., (2004), Lewis acid−base, molecular modeling, and isotopic labeling in a sophomore inorganic chemistry laboratory. *J. Chem. Educ.*, **81**, 722–724.

Nehm, R.H., Haertig, H., (2012), Human vs. computer diagnosis of students' natural selection knowledge: testing the efficacy of text analytic software. *J. Sci. Educ. Technol.*, **21**, 56–73.

O'Sullivan, T.P., Hargaden, G.C., (2014), Using structure-based organic chemistry online tutorials with automated correction for student practice and review. *J. Chem. Educ.*, **91**, 1851–1854.

Paik, S.-H., (2015), Understanding the relationship among Arrhenius, Brønsted–Lowry, and Lewis theories. *J. Chem. Educ.*, **92**, 1484–1489.

Popova, M., Bretz, S.L., (2018), Organic chemistry students' understandings of what makes a good leaving group. *J. Chem. Educ.*, **95**, 1094–1101.

Prevost, L.B., Smith, M.K., Knight, J.K., (2016), Using student writing and lexical analysis to reveal student thinking about the role of stop codons in the central dogma. *CBE-Life Sci. Educ.*, *15*, ar65: 1-13.

Respondent Fatigue, (2008), , in: Encyclopedia of Survey Research Methods. Sage Publications, Inc., 2455 Teller Road, Thousand Oaks California 91320 United States of America.

Richards-Babb, M., Curtis, R., Georgieva, Z., Penn, J.H., (2015), Student perceptions of online homework use for formative assessment of learning in organic chemistry. *J. Chem. Educ.*, **92**, 1813–1819.

Schimmel, B.J., (1983), A Meta-Analysis of Feedback to Learners in Computerized and Programmed Instruction.

Schmidt, H.-J., (1997), Students' misconceptions—Looking for a pattern. *Sci. Educ.*, **81**, 123–135.

Schmidt, H.-J., Volke, D., (2003), Shift of meaning and students' alternative concepts. *Int. J. Sci. Educ.*, **25**, 1409–1424.

Sesen, B.A., Tarhan, L., (2011), Active-learning versus teacher-centered instruction for learning acids and bases. *Res. Sci. Technol. Educ.*, **29**, 205–226.

Shaffer, A.A., (2006), Let us give Lewis acid−base theory the priority it deserves. *J. Chem. Educ.*, **83**, 1746–1749.

Sheskin, D.J., (2011), Test 20: The McNemar Test, in: Handbook of Parametric and Nonparametric Statistical Procedures. Chapman & Hall/CRC, Boca Raton, FL, pp. 835–882.

Stoyanovich, C., Gandhi, A., Flynn, A.B., (2015), Acid–base learning outcomes for students in an introductory organic chemistry course. *J. Chem. Educ.*, **92**, 220–229.

Taagepera, M., Noori, S., (2000), Mapping students' thinking patterns in learning organic chemistry by the use of Knowledge Space Theory. *J. Chem. Educ.*, **77**, 1224–1229.

Tarhan, L., Sesen, B.A., (2012), Jigsaw cooperative learning: acid–base theories. *Chem. Educ. Res. Pract.*, **13**, 307–313.

Tümay, H., (2016), Emergence, learning difficulties, and misconceptions in chemistry undergraduate students' conceptualizations of acid strength. *Sci. Educ.*, **25**, 21–46.

Ültay, N., Calik, M., (2016), A comparison of different teaching designs of 'acids and bases' subject. *Eurasia J. Math. Sci. Technol. Educ.*, **12**, 57–86.

Wang, A.Y., Thomas, M.H., (1995), Effect of keywords on long-term retention: help or hindrance? *J. Educ. Psychol.*, **87**, 468–475.

المنارة للاستشارات
www.manaraa.com

Weston, M., Haudek, K.C., Prevost, L., Urban-Lurain, M., Merrill, J., (2015), Examining the impact of question surface features on students' answers to constructed-response questions on photosynthesis. *CBE-Life Sci. Educ.*, **14**, ar19: 1-12.

للاستشارات

www.manaraa.com

Chapter 5

Analyzing explanations of substitution reactions using lexical analysis and logistic regression techniques

5.1 Note to reader

This chapter is a published manuscript in *Chemistry Education Research and Practice*. The chapter was reproduced from:

Dood, Amber J.; Dood, John C.; Cruz-Ramírez de Arellano, Daniel; Fields, Kimberly B.; Raker, Jeffrey R. (2020) Analyzing explanations of substitution reactions using lexical analysis and logistic regression techniques. *Chemistry Education Research and Practice*, **21**, 267-286.

DOI: 10.1039/C9RP00148D

with permission from the Royal Society of Chemistry. Further copyright information can be found in Appendix C.

This work has co-authors. John C. Dood is a software engineer who greatly contributed to the coding of the Python program presented in this chapter. Daniel Cruz-Ramírez de Arellano and Kimberly B. Fields are instructors of organic chemistry at the University of South Florida who allowed for data collection in their classes. Jeffrey R. Raker is the principal investigator for this project.

www.manaraa.com

5.2 Abstract

Assessments that aim to evaluate student understanding of chemical reactions and reaction mechanisms should ask students to construct written or oral explanations of mechanistic representations; students can reproduce pictorial mechanism representations with minimal understanding of the meaning of the representations. Grading such assessments is time-consuming, which is a limitation for use in large-enrollment courses and for timely feedback for students. Lexical analysis and logistic regression techniques can be used to evaluate student written responses in STEM courses. In this study, we use lexical analysis and logistic regression techniques to score a constructed-response item which aims to evaluate student explanations about what is happening in a unimolecular nucleophilic substitution (i.e., S_N1) reaction and why. We identify three levels of student explanation sophistication (i.e., descriptive only, surface level *why*, and deeper *why*), and qualitatively describe student reasoning about four main aspects of the reaction: leaving group, carbocation, nucleophile and electrophile, and acid−base proton transfer. Responses scored as Level 1 (N = 113, 11%) include only a description of what is happening in the reaction and do not address the why for any of the four aspects. Level 2 responses (N = 549, 53%) describe why the reaction is occurring at a surface level (i.e., using solely explicit features or mentioning implicit features without deeper explanation) for at least one aspect of the reaction. Level 3 responses (N = 379, 36%) explain the why at a deeper level by inferring implicit features from explicit features explained using electronic effects for at least one reaction aspect. We evaluate the predictive accuracy of two binomial logistic regression models for scoring the responses with these levels, achieving 86.9% accuracy (with the testing data set) when compared to human coding. The lexical analysis methodology and emergent scoring framework could be used as a foundation from which to develop scoring models for a broader array of reaction mechanisms.

المنارة للاستشارات
www.manaraa.com

5.3 Introduction

Research has shown that students can draw reaction mechanisms (i.e., the stepwise process of the movement of electrons over the course of a reaction) without understanding the pictorial representation of a reaction mechanism (Bhattacharyya and Bodner, 2005; Grove *et al.*, 2012; Yan and Talanquer, 2015); this body of research has relied on think-aloud interviews and constructed-response assessments to identify the observed lack of understanding. Combined, the work demonstrates that explanations are a necessary complement to mechanistic pictures in order to provide stronger evidence of understanding of reaction mechanisms (c.f., Becker *et al.*, 2016; Crandell *et al.*, 2018). Despite their importance, use of constructed-response items during instruction and on formative assessments is limited due to time constraints of grading written responses and lack of automated scoring tools.

Lexical analysis is a solution to scoring written responses to constructed-response items including written explanations of reaction mechanisms. Through lexical analysis and resultant predictive models, computers are trained to code written responses to constructed-response assessments for themes and overall response levels of sophistication or correctness. Such models have been used in STEM education to formatively assess student understanding of phenomena such as acid−base chemistry (e.g., Haudek *et al.*, 2012; Dood *et al.*, 2018), the central dogma of biology (e.g., Prevost *et al.*, 2016), and thermodynamics (e.g., Prevost *et al.*, 2012). Results of computer-scored written responses provide data from which educators can tailor future instruction including remediation or forgo further discussion of topics that a majority of students already understand.

In this study, we describe the use of lexical analysis and the application of logistic regression model techniques to the scoring of a constructed-response assessment item aimed to evaluate student understanding of a substitution reaction, one of several reaction types taught in introductory organic chemistry courses. We identify three levels of explanation sophistication in student responses to what happens in the reaction and

المنارة للاستشارات
www.manaraa.com

why the chemical reaction occurs mechanistically. We evaluate the predictive accuracy of a pair of binomial logistic regression models that, when combined, produce our three-level scoring of assessment responses. Lastly, we offer methods for incorporating the assessment item and resultant computer-scoring model as a formative assessment in chemistry instruction.

5.4 Student understanding of organic chemistry mechanisms

Organic chemistry mechanisms are represented by curved arrows that signify the movement of electrons from areas of high electron density to areas of low electron density (Bhattacharyya, 2013; Kermack and Robinson, 1922). Practicing organic chemists use mechanisms as tools to predict the products of reactions. Research has shown that students do not view mechanisms as tools (e.g., Bhattacharyya and Bodner, 2005; Grove *et al.*, 2012). Instead, students memorize the pictorial representations without connecting chemical theory to the lines, letters, and symbols; they then reproduce the mechanistic pictures to earn points on assessments, where, for example, understanding of mechanisms is frequently assessed solely through reproduction of pictorial representations. Grove *et al.* (2012) found that students drew the chemical structures of mechanistic steps from memory and then added in arrows at the end, as if decorating the structures with arrows, instead of using mechanisms as a communicative and predictive tool. Bhattacharyya and Bodner (2005), as well, found that students were able to successfully reproduce organic chemistry mechanisms, but did not understand the chemical concepts behind the mechanisms.

Student performance in organic chemistry courses has been correlated with type of mechanistic explanation (Cooper *et al.*, 2016; Crandell *et al.*, 2018; Dood *et al.*, 2018). Cooper *et al.* (2016) and Crandell *et al.* (2018) found that students who provided causal mechanistic reasoning in response to an assessment item were more likely to successfully produce the mechanistic representation. A similar constructed-response item was used in our previous work (Dood *et al.*, 2018) which found that Lewis acid−base model use in

المنارة للاستشارات
www.manaraa.com

mechanistic explanations of an acid−base proton-transfer reactions was associated with higher performance on acid−base examination items and overall exam performance.

Explanations and descriptions of pictorial representations have been shown to uncover understanding and misunderstanding of organic chemistry and related topics where merely asking students to produce representations fails to do so (Anzovino and Bretz, 2015, 2016; Bhattacharyya and Bodner, 2005; Bhattacharyya and Harris, 2018; Cooper *et al.*, 2016; Cruz-Ramírez de Arellano and Towns, 2014; Dood *et al.*, 2018; Grove *et al.*, 2012; Popova and Bretz, 2018; Strickland *et al.*, 2010). Even entry-level chemistry graduate students who are intending to study organic chemistry lacked the understanding of mechanisms that instructors expect from undergraduate organic chemistry students (Strickland *et al.*, 2010). Constructed-response assessments provide a means to evaluate understanding of reaction mechanisms via student explanations about provided pictorial drawings or coupled with assessments that have students generate their own pictorial drawings.

5.5 Reasoning in chemistry

Students and instructors use multiple types of reasoning when considering chemistry-related concepts (e.g., teleological, anthropomorphic, mechanistic, causal, causal mechanistic). Teleological reasoning implies purpose based on a certain goal (Talanquer, 2007; Tamir and Zohar, 1991; Wright, 1972, 1976). For example, a common example of teleological reasoning is that atoms react the way they do in order to fulfill the octet rule (Talanquer, 2007). Caspari *et al.* (2018b) asked students to propose mechanistic arrows for organic chemistry reactions and explain why they proposed each step. Teleological reasoning was observed, with some students claiming that the reason for the occurrence of a mechanistic step was in order to make the next step possible. Similarly, students in a study conducted by Bhattacharyya and Bodner (2005) reasoned that mechanistic steps occurred in order to get to the product. Anthropomorphic reasoning, frequently

المنارة للاستشارات
www.manaraa.com

used in conjunction with teleological reasoning, attributes human-like characteristics to non-human entities (e.g., "the atom *wants* to have a full octet"). Anthropomorphic reasoning can be constructive in cases where teachers and students are aware that anthropomorphism is being used metaphorically (Lemke, 1990; Taber and Watts, 1996). However, anthropomorphic and teleological reasoning become problematic when students think entities such as atoms and electrons actually *want* and *need*. Talanquer (2013) stated that anthropomorphic and teleological explanations provide people with a false sense of understanding, leading them not to seek out deeper understanding of phenomena because they have confused their superficial understanding with deeper understanding. Talanquer (2013) also called teleological explanations "a cognitively cheap way of satisfying a need for explanation without having to engage with more complex mechanistic reasoning" (p. 1423); Talanquer (2013) argued that students in chemistry classrooms should be given the opportunity to develop and apply mechanistic and causal explanations of chemical phenomena.

In order to help students develop reasoning beyond teleological reasoning, instruction should focus on *how* and *why* chemical process are occurring at mechanistic and causal levels. According to Talanquer (2018), mechanistic reasoning "invoke[s] the existence of entities (e.g., atoms, molecules) whose properties, interactions, activities, and organization are responsible for the behaviors we observe" (p. 1906). Teaching mechanistic reasoning in the classroom is a critical part of instruction, as it describes "how the particular components of a system give rise to its behavior" (Russ *et al.*, 2008, p. 504). In addition to mechanistic reasoning, instructional activities should encourage students develop reasoning that explains the process of cause and effect (i.e., causal reasoning; Abrams and Southerland, 2001; Koslowski, 1996; Schauble, 1996; Sperber *et al.*, 1996; Zimmerman, 2000). In the context of chemistry, Yan and Talanquer (2015) define chemical mechanism as *how* reactions happen and chemical causality as *why* reactions happen, arguing that chemical causality is necessary to demonstrate full understanding of reactivity. An example from a framework presented by Crandell *et al.* (2018) describes causal reasoning/explanations in the context of acid−base reactions as discussing the electrostatic interaction between species, while mechanistic reason-

المنارة للاستشارات
www.manaraa.com

ing/explanations only describe the movement of electrons. Causal mechanistic reasoning combines aspects of causal and mechanistic reasoning. In the context of organic chemistry, Crandell *et al.* (2018) defined causal mechanistic reasoning as “a type of explanation of a phenomenon that identifies the causal factors and the activities of the underlying entities (electrons) to provide a stepwise account of the phenomenon from start to finish” (p. 214). Crandell *et al.* (2018) argued that asking students to engage in causal mechanistic reasoning is beneficial, as the act “requires that students reflect on and connect the sequence of events underlying a phenomenon and the causal drivers involved” (p. 215), therefore facilitating learning. The idea that causal mechanistic reasoning should promote learning is supported by the results of Cooper *et al.* (2016) who found that students who engaged in causal mechanistic reasoning were more likely to successfully produce mechanistic arrows. Therefore, instruction in organic chemistry should emphasize the mechanistic and causal reasoning behind chemical processes, including asking students to explain such process in assessment contexts.

5.5.1 Reasoning about substitution reactions

Studies have explored student understanding of substitution reactions and aspects of substitution reactions (Anzovino and Bretz, 2016, 2015; Bodé *et al.*, 2019; Caspari *et al.*, 2018a; Cruz-Ramírez de Arellano and Towns, 2014; Popova and Bretz, 2018). An interview study by Cruz-Ramírez de Arellano and Towns (2014) considered student understanding of alkyl halide reactions similar to the S_N1 reaction presented in this study. Students were asked to solve organic chemistry mechanism problems requiring them to predict the products of alkyl halide reactions. An “expert argumentation scheme” included warrants such as: classifying methanol as a weak nucleophile; knowing that chlorine will act as a leaving group and create a carbocation intermediate; and that “oxygen in the substitution product will lose its proton through acid−base chemistry with another molecule of methanol” (p. 505).

www.manaraa.com

The results of these studies indicate that student reasoning about aspects of substitution reactions remains at a surface level. Students memorize which leaving groups are *good*, but are unable to describe why in a meaningful way, invoking the octet rule and charge to size ratio to reason about leaving group stability, and electronegativity and electron pulling as reasons for halides being good leaving groups (Caspari *et al.*, 2018a; Popova and Bretz, 2018). Understanding of carbocation stability is limited to the number of substituents of the carbocation, again indicating reasoning that is limited to explicit features (Bodé *et al.*, 2019; Caspari *et al.*, 2018a). Studies on students' understanding of nucleophiles and electrophiles have shown evidence of heavy reliance on explicit reaction features, with students using structural cues such as charges and mechanistic arrows to identify nucleophiles and electrophiles (Anzovino and Bretz, 2016, 2015).

5.5.2 Levels of reasoning in organic chemistry

Levels of reasoning have been used to describe student understanding of reaction mechanisms (Bodé *et al.*, 2019; Caspari *et al.*, 2018a; Sevian and Talanquer, 2014). Caspari *et al.* (2018a) asked students to explain which of two mechanistic steps had the lowest activation energy. Based on student explanations, Caspari *et al.* (2018a) defined levels of complexity of relations that included low complexity, middle complexity, and high complexity (Table 5.1). Bodé *et al.* (2019) presented students with mechanisms and reaction coordinate diagrams for two similar reactions and asked which reaction was most likely to proceed and why. Using the Chemical Thinking Learning Progression (CTLP) developed by Sevian and Talanquer (2014) as a framework, Bodé *et al.* (2019) scored student responses as descriptive, relational, linear causal, and multicomponent causal (Table 5.2). Most student responses (63%) fell into the linear causal category. Although multicomponent causal was considered ideal, no student responses fell into that category. Caspari *et al.* (2018a) and Bodé *et al.* (2019) included the theme of explicit versus implicit features in their scoring schemes; ideally, explanations would include implicit features inferred from explicit features. The

المنارة للاستشارات
www.manaraa.com

CTLP, as used by Bodé *et al.* (2019), considered non-causal versus causal reasoning, with two lower levels for non-causal reasoning and two more advanced levels for causal reasoning. These studies elicited and described levels of students' chemical reasoning using constructed-response items. As the construction of explanations in science is an important tool for learning, other studies have specifically addressed the nature of assessment prompts to promote and elicit student reasoning.

Table 5.1 Scoring scheme used by Caspari *et al.* (2018a) to analyze student understanding of reaction mechanisms

Relations with low complexity: **explicit structural differences** or **non-electronic effects** used as a cause for change.
Relations with middle complexity: **implicit structural properties** or **non-electronic effects** used as a cause for change.
Relations with high complexity: **implicit structural differences** used to describe **electronic effects** on change.

Table 5.2 Scoring scheme used by Bode *et al.* (2019) which was adapted from the CLTP developed by Sevian and Talanquer (2014)

Descriptive: **describing properties** of reaction materials, **explicit (surface) features** of problem described.
Relational: **explicit and implicit properties** discussed; connections made but reasoning does not get to the why.
Linear causal: **cause-and-effect relationships** (i.e., the why) are discussed for **single variables** using explicit and implicit properties of the reaction.
Multicomponent causal: **multiple variables** considered using explicit and implicit features of the reaction to describe **cause-and-effect relationships**.

5.6 Constructed-response items to elicit chemical reasoning

Purposefully developed constructed-response items can elicit and promote student understanding of scientific representations and concepts. *The Framework for K-12 Science Education* names constructing explanations and engaging in argument from evidence as two of eight practices for science classrooms (National Research Council, 2011) that help develop understanding. The Framework states that scientific explanations "explain observed relationships between variables and describe the mechanisms that support cause and effect inferences about them" (p. 67). Becker *et al.* (2016) stated that "engaging students in the construction of scientific explanations affords them an opportunity to engage (in a scaffolded form) in the

المنارة للاستشارات
www.manaraa.com

practices of science, that is, the activities that scientists engage in as they investigate the world" (p. 1714). When the proper scaffolding is provided, constructed-response items can be used to elicit student reasoning in chemistry and to promote student construction of knowledge through participation in the development of scientific explanations.

Constructed-response items are useful for eliciting student explanations when properly developed. Recommendations from previous work to better elicit student explanations included providing students with multiple representations (Becker *et al.*, 2016; Crandell *et al.*, 2018) and specifically asking students both *what* is happening and *why* (Cooper *et al.*, 2016). Though useful for eliciting and developing explanations, constructed-response items are onerous to incorporate in courses, as time is required for an instructor to read and score responses. The delay in feedback due to grading is a barrier to the utility of constructed-response items for formative assessment purposes. Constructed-response items could be more useful if incorporated into automated assessment systems with computer-assisted scoring where immediate feedback is possible. Instructors would benefit from the amassed feedback provided as this feedback could be used to develop new instructional strategies or just-in-time teaching moments in the classroom. Students, as well, could benefit from immediate feedback received about their written responses. Lexical analysis techniques coupled with predictive regression models provide a means for development of the type of computer-assisted scoring necessary to achieve the utility of constructed-response items.

5.7 Lexical analysis of written responses

Lexical analysis is a technique to evaluate textual responses such as those obtained through constructed-response items. This technique codes common words and phrases used in text data; codes are then used to construct predictive models (e.g., binomial logistic regression). Such models have been used to predict correctness, conceptual understanding, and levels of understanding for a number of constructed-response items

المنارة للاستشارات
www.manaraa.com

used in chemistry and biology instruction (Dood *et al.*, 2018; Haudek *et al.*, 2012; Kaplan *et al.*, 2014; Moharreri *et al.*, 2014; Prevost *et al.*, 2016; Prevost *et al.*, 2013; Prevost *et al.*, 2012; Shen *et al.*, 2014).

Lexical analysis begins with reviewing text-based responses for common words and phrases. These are then combined into broader categories that, in the case of use with constructed-response items in chemistry, may indicate a concept or topic being used or not used by a student within a written response. For example, common words and phrases used to describe differences in electron density in a reactant molecule could be combined into a category called *partial charges*; any response to the assessment item that uses any of the words or phrases associated with this category would be marked accordingly: *1* means included in the response and *0* means not included in the response. After categories have been developed for a data set, those categories are used to develop a regression model that predicts an overall score for each response based on categories represented in that response. In previous work, we coded responses for *use* or *non-use* of a Lewis acid−base model by students when describing the mechanism of a single-proton transfer reaction (Dood *et al.*, 2018). We used a binomial logistic regression model with categories (such as *partial charges* described above) as predictors of Lewis acid−base model use, a binary variable. Logistic regression models can be developed for binary, ordinal, and multinomial models based on the desired overall scoring scheme of the constructed-response item. As an example of a contrasting model, a multinomial logistic regression model was used by Prevost *et al.* (2016), scoring student responses as: correct, incomplete/irrelevant, or incorrect.

A high level of predictive accuracy (greater than 85% agreement with human coding) is expected for determining the utility of the computer-assisted scoring of assessment items. A large number of diverse responses is necessary to develop a model with broad predictive ability. While no established guidelines exist for minimum or ideal sample size, the Automated Analysis of Constructed Response (AACR) and associated studies typically collect more than 800 responses to build and evaluate scoring models; such a sample size provides a sufficient number of responses to divide the data into a training set (70%) from

المنارة للاستشارات
www.manaraa.com

which the model is built and refined and a testing set (30%) from which the model is evaluated. Use in large enrollment courses is touted as a key utility of the computer-assisted scoring of constructed-response items (Prevost *et al.*, 2013); typical testing sets (30% of 800) equate to approximately 240 responses, a likely size of a large enrollment course.

Examples of lexical analysis and predictive regression models being used to analyze constructed-response assessments span only a small number of scientific disciplines: biology (Haudek *et al.*, 2012; Moharreri *et al.*, 2014; Prevost *et al.*, 2016, 2012), chemistry (Dood *et al.*, 2018), and statistics (Kaplan *et al.*, 2014). A collection of constructed-response items can be found in the AACR library (beyondmultiplechoice.org). The project began in biology, and has now branched out to biochemistry, chemistry, and statistics topics; however, items topics are often central and relevant to biology. In this study, we analyze student explanations of an S_N1 reaction mechanism and develop a predictive regression model to evaluate student understanding of the S_N1 reaction.

5.8 Theoretical foundation and research questions

This study embraces an emergent process for identifying and predicting classifications of students' mechanism explanations, and presents levels of sophistication of student explanations which arose from the data set. Our analysis is framed in the current understanding in the literature of student explanations and descriptions of mechanisms. Given this understanding, the goal of this study is to accurately code students' written responses to a constructed-response item using a predictive logistic regression model that has potential to be used in classes with large enrollment. We address two key questions:

1. How do students respond to a prompt asking for a written response about what is happening and why for an S_N1 (i.e., unimolecular substitution) reaction mechanism?

www.manaraa.com

2. Does lexical analysis lead to a logistic regression model for predicting level of explanation sophistication in responses to a constructed-response formative assessment item on explaining an S_N1 reaction mechanism?

5.9 Methods

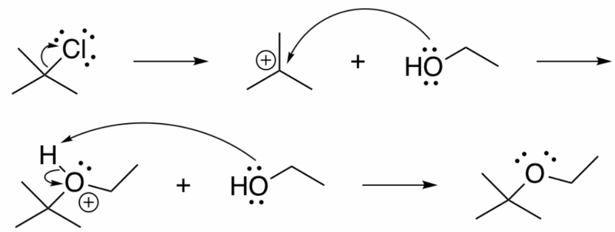
This work was conducted under application Pro#00028802, "Comprehensive evaluation of the University of South Florida's undergraduate and graduate chemistry curricula", as reviewed and approved on December 13, 2016, by the University of South Florida's Institutional Review Board.

5.9.1 Development of constructed-response item

The constructed-response item (see Figure 5.1) used in this study mirrors an item developed by Cooper *et al.* (2016) which asked students in a transformed general chemistry curriculum what is happening in an acid−base proton-transfer mechanism at a molecular level and why. Cooper *et al.* used an iterative, research-based process to develop the item. In previous work, we adapted the item for use in our organic chemistry curriculum and for scoring using lexical analysis and logistic regression (Dood *et al.*, 2018). The constructed-response item in this study uses wording and symbolism that mirror the prompt used in our previous work; the key difference is this study explores a multistep S_N1 reaction mechanism rather than a single-step proton-transfer reaction. Additionally, the predictive model in our previous work coded responses for use or non-use of the Lewis acid−base model while the predictive model in this study scores responses for overall level of explanation sophistication. To broaden the utility of the item, the leaving group was varied (i.e., bromide, chloride, and iodide). The version of the prompt with chloride as the leaving group is depicted in Figure 5.1.

www.manaraa.com

Consider the mechanism below for the S_N1 reaction between t-butyl chloride and ethanol to form ethyl t-butyl ether.



A. Describe in full detail *what* you think is happening on the molecular level for this reaction. Be sure to discuss the role of each reactant.

B. Using a molecular level explanation, please explain *why* this reaction occurs. Be sure to discuss why the reactants form the products shown.

Figure 5.1. Constructed-response items given to students with chloride as the leaving group. Additional iterations of the prompt included bromide and iodide as leaving groups.

Students were provided with separate response boxes for parts A and B of the prompt, as Cooper *et al.* (2016) found that including both what and why in separate parts better elicited student reasoning. However, similar to Cooper *et al.*, responses to parts A and B were combined for analysis due to many students being unable to differentiate between *what* and *why*.

5.9.2 Data collection

Data were collected during the Fall 2017 and Spring 2018 semesters in the first semester of a two-semester postsecondary organic chemistry sequence with three different instructors at a large, public university in the southeast United States. The item was given as an extra credit opportunity via Qualtrics. The iteration of the item with bromide as a leaving group (N = 733) was collected during the Fall 2017 semester, and the iterations with chloride (N = 155) and iodide (N = 153) as leaving groups were collected during the Spring 2018 semester.

المنارة للاستشارات
www.manaraa.com

5.9.3 Development of scoring scheme

Responses were first analyzed by author AJD to inductively develop a scoring scheme via an exploratory and iterative process (c.f., constant comparative analysis; Glaser and Strauss, 1967; Strauss and Corbin, 1990), allowing the scoring scheme to emerge from the data. Through this process, author AJD noted different levels of complexity in students' explanations based on four main aspects of the presented reaction: leaving group, carbocation, nucleophile and electrophile, and proton transfer. Different explanations types were classified into categories through discussions between authors AJD and JRR, during which categories were refined and inclusion criteria were established. Author AJD then independently applied this initial scheme to all responses. The resulting preliminary categories described the level of complexity of student explanations: description only (i.e., response describes *what* is occurring but not *why* the reaction occurs), surface level reasoning (i.e., reasoning about why the reaction occurs at a surface level), and deeper reasoning (i.e., reasoning about why the reaction occurs at a deeper level). To test the fidelity of the reasoning categories, author JRR performed an interrater check by randomly selecting responses across the three categories and applying the categories; discrepancies were discussed between authors AJD and JRR until agreement was reached. Minor revisions to the category descriptions were made and author AJD applied the final scheme to the remaining responses. After the sorting process was complete, it was determined that the categories represented levels of explanation sophistication. Therefore, we refer to the categories as Level 1, Level 2, and Level 3.

5.9.4 Development of logistic regression model

Model development and evaluation was conducted using Python, an open-source scripting language (code available at https://osf.io/muw3r/?view_only=7ceacdb889704083b098a4239235e2dd), and a lexical analysis framework built using SPSS Modeler. Logistic regression analysis code for Python was adapted from Pedregosa *et al.* (2011). The model development process is outlined in Figure 5.2. Combined re-

المنارة للاستشارات
www.manaraa.com

sponses to part A and part B of the assessment item were randomly partitioned into a training set (N = 728, 70%) and a testing set (N = 313, 30%). Training data were used to build and revise the logistic regression model; testing data were used after the model was built to validate the model. The foundation of the predictive model is a hierarchical coding scheme using common words and phrases to generate categories (i.e., predictor variables). A pair of binomial logistic regression models were used to predict the human-determined score.

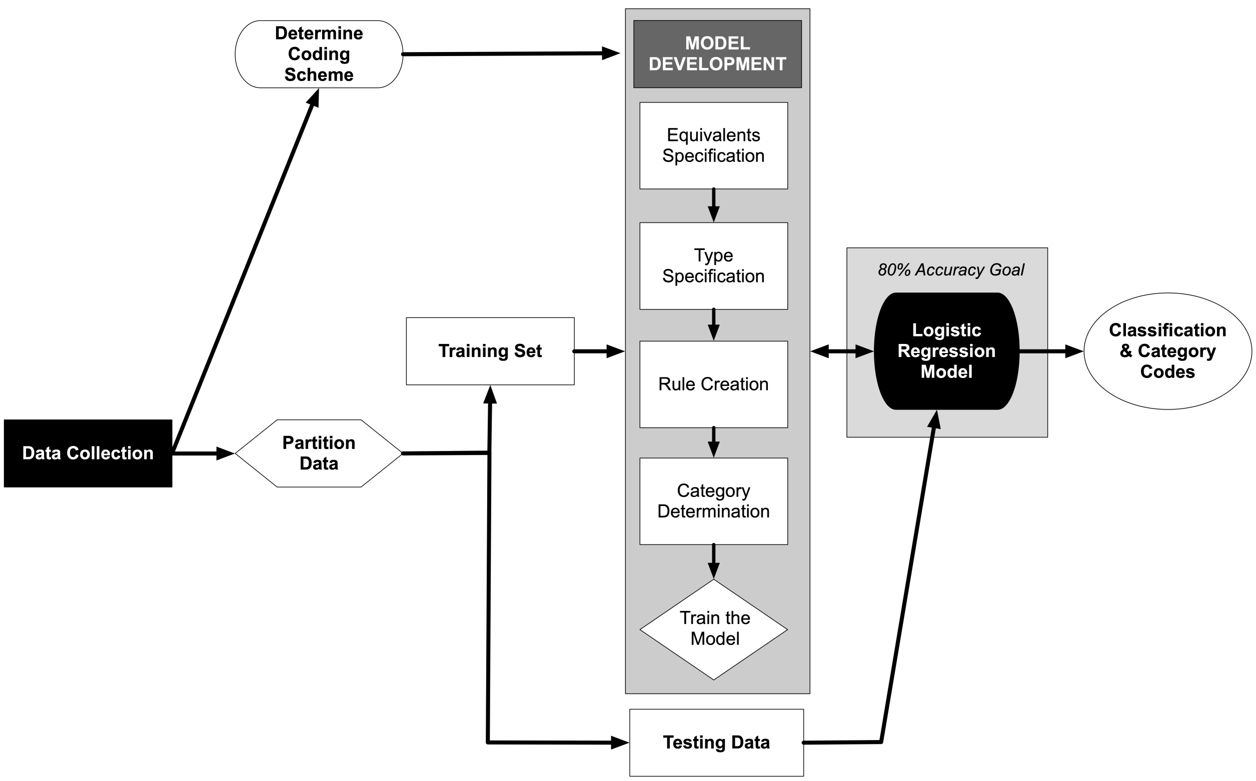


Figure 5.2. The process of creating a logistic regression predictive model from constructed-response item to model development.

The first step in model development was to generate four coding algorithms that prepared and analyzed the text responses for inclusion in the predictive model. The first two algorithms replaced words or phrases in the responses to adjust for misspellings, equivalents, or synonyms. The last two algorithms con-

www.manaraa.com

sidered combinations of words or phrases in the responses and generated categorical codes used as predictor variables in the regression models.

Equivalents

Equivalents are words or phrases that read as exactly the same, common misspellings, and different ways to write something. For example, consider *bromine*. This term can include Br, bromide, and Br–, as students are referring to the same thing when using these in the context of our assessment item. The term *bromine* also includes "bromin" and "bromean" as common misspellings. The equivalents algorithm cleans responses by replacing all equivalents to the same term; in the case of our example, all of the words in the response that are equivalent to *bromine* (e.g., "bromin" or Br) will be replaced with the term *bromine*.

Types

Types are groups of words that are synonyms. For example, a type called *halogen* could include: bromine, chlorine, iodine, and fluorine. In the context of our assessment item, these words are synonyms because three different halogens are used as the leaving group in the three iterations of the assessment item. As with the equivalents algorithm, data are cleaned by replacing words or phrases with a synonymous word or phrases (e.g., *halogen*).

Rules

Rules are patterns of text made up of words, phrases, equivalents, or types. A rule consists of two or more words, equivalents, or types that appear in the text in sequence within six or less words of each other. For example, the phrase "the alcohol is **attracted** to the **positively** charged carbocation" contains words or phrases of the type *attract* and the type *positive* (see bolded words). Though the two words are not directly next to each other, a rule that codes for the *attract* and *positive* types would apply to this phrase.

المنارة للاستشارات
www.manaraa.com

The order of the words, equivalents, or types is important. Thus, separate rules must be created to account for discrepancies in word order, as respondents may use different wordings to mean the same thing (e.g., "the bond was formed" vs. "it formed a bond"). The rules *bond forms* and *forms bond* were thus both necessary. The rule *bond forms* must contain a word from the type *bond* and then a word from the type *form*. The rule *forms bond* must contain a word from the type *form* and then a word from the type *bond*. The necessary order of terms has the potential to distinguish between understanding and misunderstanding in cases when use of the terms is parallel with correct and incorrect application of a scientific concept or principle, though the data set in this study was not scored for correctness.

Categories

A category contains specific words, phrases, rules, or types found in responses; a category may also be designed to denote that a response does not contain particular words, phrases, rules, or types. Any given response can be coded with all or none of the categories, with most responses being coded with multiple categories (an average of 9 categories per response out of 33 possible categories for responses in this data set). A numeric code of *1* is given to responses that include a given category or a numeric code of *0* is given to responses that do not include the given category.

An example category is *bond forming*; this category includes any responses that are coded as use of the rules *bond forms* and *forms bond*, as well as the phrase"new bond". Another example category is *absence of explanation*; this category includes responses that omit words from types that would indicate the student has addressed why the reaction is occurring.

Logistic regression model

Category codes (i.e., *1* or *0*) are then used as predictor variables in logistic regression models. In this study, we used a pair of binomial logistic regression models with human-scoring as the outcome variable

المنارة للاستشارات
www.manaraa.com

due to higher accuracy in this context over an ordinal model. The model development process used only the training data set. Accuracy was determined as the percentage of correct scoring assignments of the responses by the computer as compared to human scores. An iterative process of refining equivalents, types, rules, and category coding algorithms was conducted with the training dataset until model accuracy was maximized. Published models typically include accuracy of at least 70%, although the desired level of accuracy is typically >85%. Refinement of algorithms was achieved through consideration of how algorithms led to false positive and false negative predicted scores. When accuracy was maximized, the resultant model was applied to the testing set.

5.10 Results and discussion

5.10.1 RQ1: How do students respond to a prompt asking for a written response about what is happening and why for an S_N1 reaction mechanism?

Analysis of responses to the constructed-response item resulted in a tri-level score based on complexity of explanation used by students when writing about what is happening and why for an S_N1 reaction mechanism (see Table 5.3 for representative responses). A Level 1 response was limited to a description of what is happening in the reaction and did not address why the reaction occurs. A Level 2 response included a surface level explanation about why the reaction occurs. That is, the explanation was related only to explicit features of the reaction or mentioned implicit features, but with no evidence of deeper reasoning. For example, if an explanation claimed bromide is a good leaving group due to its size with no other explanation, it is impossible to tell whether the explanation is based on sound chemical reasoning or if the explanation stems from memorizing a pattern of atomic size. A Level 3 response included a deeper level explanation of why the reaction occurs (i.e., reasoning related to implicit features of the reaction that have been inferred from the explicit features of the reaction and explained using electronic effects). The three levels are hierar-

www.manaraa.com

chical in that a Level 3 response could include similar phrases to a Level 1 or Level 2 response as to what is happening in the reaction; however, the Level 3 response would provide a deeper explanation for why the reaction is happening that is not present in a Level 1 or Level 2 response. In the complete human-coded data set, there were 113 responses coded as Level 1, 549 responses coded as Level 2, and 379 responses coded as Level 3.

Table 5.3 Scheme used to score the constructed-response item.

Overall score and description	Example
Level 1. Response only describes *what* is happening in the reaction; the response does not address *why* the reaction is occurring.	"A nucleophilic attack is occurring. The Br is the leaving group and this happens in two steps. The O is attacking the cation and then grabbing the hydrogen."
Level 2. Response describes *why* the reaction is occurring at a **surface level**; explanations either include only explicit reaction features or mention implicit features with no further explanation. Example reasons: stability and leaving group ability.	"This reaction occurs because bromide is a *good leaving group* that generally wants to break a bond in order for a nucleophile such as water to come in and *balance the charge*."
Level 3. Response describes *why* the reaction is occurring at a **deeper level**; explanation includes implicit features of the reaction that have been inferred from explicit features; explains electronic effects. Example reasons: electron density, electronegativity, and partial charges.	"Since iodine is such a good leaving group *due to its large electron shell and thus polarizability*, the bond between the alpha carbon and iodine is polar. Once the iodine leaves, the alpha carbon forms a tertiary carbocation. *Due to hyperconjugation and inductive effects*, this is as stable as it gets in terms of carbocations. The polar secondary alcohol (*with oxygen having a partial negative charge*) is *attracted to the full positive charge* of the tertiary carbocation and therefore attacks it."

Though overall levels were assigned holistically, responses were noted to include the four components of S_N1 reactions of which explanation could be considered in levels: leaving group, carbocation, nucleophile and electrophile, and proton transfer. Responses varied in the number of components that were addressed from none to all four. Responses that addressed more components generally received a higher-level score; however, a response could address all four components but provide no explanation for why parts of the reaction act the way they do and thus still receive a Level 1 score. In addition, it was not essential that explanation be offered for each component explicitly to receive a Level 2 or Level 3 score. For example, a respondent could explain that the leaving group bromine will be a good leaving group because of its large atomic size (this is a memorized trend with no evidence of deeper reasoning and thus an example of Level

المنارة للاستشارات
www.manaraa.com

2 explanation), but only describe the other parts of the reaction at a Level 1. The overall Level assignment would be Level 2 because of the inclusion of the why at a surface level for at least one component of the reaction.

Differences in how respondents addressed the four components (i.e., leaving group, carbocation, nucleophile and electrophile, and proton transfer) provide more information about reasoning used at each scoring Level. An example response developed by the authors and considered expert-level with bromide as the leaving group is provided below with specific deeper reasoning bolded for emphasis:

An S_N1 *reaction is occurring. In the first step, bromide acts as a leaving group and the bond between bromine and carbon is broken.* ***Bromine's large atomic radius allows for a stable bromide ion due to delocalization of the negative charge.*** *The positively-charged tertiary carbocation is also* ***stable because of shared electron density from surrounding methyl groups, which act as electron donors.*** *In the second step, ethanol acts as a nucleophile and the carbocation acts as an electrophile. The electrostatic attraction between the carbocation and the nucleophile is greater than that of the carbocation and bromide; thus,* ***electron-deficient carbocation and the partial negative charge of the electron-rich oxygen combine to form a new bond.*** *In the final step of the reaction, the positively-charged intermediate is neutralized. Ethanol is* ***electron-rich and able to act as a base and a bond can be formed between the ethanol-oxygen and the hydrogen atom on the product. The acidic hydrogen is electron-poor due to the positively-charged, electronegative oxygen unequally sharing electron density with it.*** *The hydrogen-oxygen bond is broken and the electrons that formed the bond become a lone pair on oxygen, resulting in a neutral final product.*

Although this response is considered expert-level for our prompt, students are not necessarily expected to be able to address all aspects of the response at this level to be successful in the course. Based on the

المنارة للاستشارات
www.manaraa.com

expectations of the course, students should be able to explain the four components of the reaction at Level 2.

Leaving group

The first step of the S_N1 reaction is the bond breaking between leaving group and substrate (i.e., starting material). Level 1 responses included descriptive statements such as “the Br is the leaving group, so the bond between Br and the carbon first breaks in the rate limiting step”. There is a description of what is happening in the response; however, the response lacks explanation of why the step is occurring. Level 2 responses included simple rationales. For example, “The Br group leaves because it is a good leaving group”. While the statement “good leaving group” is superficial, the evaluative statement suggests the foundation for a rationale for the step. However, this response could be mere evidence of a memorized fact rather than deeper level explanation. Of our 1,041 total responses, there were 339 instances of the phrase “good leaving group”, many of which were not accompanied by further explanation of why the leaving group was good. Students may have memorized which leaving groups are relatively good and used explicit features in the pictorial representation (i.e., halide as the leaving group) to reason that the reaction was able to occur due to the goodness of the leaving group. More in-depth (Level 3) reasoning would be ”bromine leaves because it is large and able to accommodate a negative charge”. We have scored this at a Level 3 based on similar higher-level explanation designations noted in previous studies showing that invoking size and its relationship to charge stabilization when explaining leaving group ability trends towards more causal explanations (Caspari *et al.*, 2018a; Popova and Bretz, 2018).

Most students in our study articulated that the halogen leaves because it is a good leaving group. Organic chemistry students often memorize relative leaving group abilities and use these memorized trends to determine the best leaving group without understanding why that leaving group is the best; Popova and Bretz (2018) made this same observation in their work on student understanding of what makes a good leav-

المنارة للاستشارات
www.manaraa.com

ing group. Though the memorized trend can be successful (i.e., a high grade), regurgitation of the trend is not evidence of chemical understanding. When assessments lack an explanatory component, memorization strategies will dominate student approaches to studying; when more constructed-response based assessments ask for explanations that require the *why*, students will need to utilize appropriate study strategies to develop the deep understanding necessary to infer implicit features from explicit features and provide causal rationales for observed phenomena such as leaving group trends. Caspari *et al.* (2018a) make a similar recommendation, calling for instructors to require extensive explanations from students and make underlying causal components the main focus of examination questions. It is important to note, though, that as assessments shift toward more constructed-response explanations, teaching approaches must shift as well. Instructors should build opportunities for students to learning how and to practice constructing explanations of phenomena causally during class time and on low-stakes assessments.

Carbocation

Students also described the formation of a carbocation intermediate in the reaction. Level 1 responses noted that a carbocation was formed; some also labeled the carbocation as tertiary. Some responses showed evidence of teleological reasoning, saying that the leaving group left because a stable tertiary carbocation would be formed; these teleological explanations were also scored at Level 1. Level 2 responses noted that the 'tertiary' carbocation was stable, sometimes noting this stability was due to the number of substituents but not elaborating on why the number of substituents matters for stability. This is evidence of a memorized pattern that more highly substituted carbocations are more stable. Level 3 responses stated that the carbocation formed is stabilized by the three electron-donating methyl groups adjacent to the cation; few responses included this level of sophistication. At Level 3, electron density should serve as the foundation for rationalizing the viability of the intermediate carbocation.

المنارة للاستشارات
www.manaraa.com

As with our critique of how students rationalized leaving groups, the lack of observed deeper level explanations using implicit features for the viability of the carbocation intermediate can also potentially be attributed to instructional and assessment strategies used in many postsecondary organic chemistry courses, including the course taken by the students who participated in this study. Though not the case in all introductory organic chemistry courses, many courses include assessments which ask students to label given carbocations as primary, secondary, tertiary, etc. Such activities may unintentionally emphasize the importance of assigning surface-level labels based on explicit features and memorizing patterns. Assessments commonly ask students to rank carbocations from least to most stable; this task can be completed by assigning the relative stability through regurgitation of a memorized trend list, however complicated the trend may get by the end of the yearlong course. The ability to correctly answer such questions solely using memorized trends can encourage students to avoid the task of developing a rationale for why the answer is correct. Using the degree of substitution on the carbocation to assert stability is a step toward developing causal reasoning for its formation; however, this surface feature-based heuristic for rationalizing stability can become problematic when multiple structural features need to be considered to determine the viability of the carbocation intermediate (e.g., allylic or benzylic carbocations). To encourage the development of deeper reasoning, students should be asked to provide a rationale for *why* a particular carbocation intermediate is more or less stable than a comparison carbocation intermediate. Caspari *et al.* (2018a) were able to elicit student explanations about carbocation stability beyond the number of substituents as deeper explanations were required to successfully answer their prompt. It is possible that our students would have invoked electron density more frequently had we provided them with an assessment where there were other obvious factors impacting carbocation stability, such as the neighboring carbonyl group present in the prompt used by Caspari e*t al.* (2018a). Students should be provided with assessments that require going beyond the number of substituents to determine carbocation stability to encourage development of deeper explanations. Items containing scaffolded questions that direct and guide students to use electron density in their responses

المنارة للاستشارات
www.manaraa.com

communicate the importance of rationalizing carbocation stability beyond a surface level. Students should also be asked to solve more complicated problems with answers that cannot be attained using only surface level reasoning to encourage them to develop the deeper level of explanation required to solve the problem at hand.

Nucleophile and electrophile

The second step of the reaction is the formation of a bond between an electrophile (i.e., the carbocation) and a nucleophile (i.e., an ethanol molecule in our assessment prompt). Responses were scored at Level 1 when the response only included labeling of the reagent and intermediate with the terms nucleophile and electrophile, respectively, and describing that a bond formed between the two. Responses that included rationale such as ethanol and the tertiary carbocation *want* to form a bond due to the positive charge of the carbocation were scored at Level 2. This reasoning is both anthropomorphic and teleological but may be a step in the right direction to developing electronic reasoning. Finally, responses that provided a rationale including electron density or partial charges, implicit features of the reaction, were scored at Level 3.

Students were much more likely to articulate the presence of a nucleophile in their responses than to articulate the presence of an electrophile. Of our 1,041 response, there were 611 instances of the term "nucleophile" and only 33 instances of the term "electrophile". This overemphasis on nucleophiles over electrophiles is in line with the work of Anzovino and Bretz (2016) who partly attributed increased familiarity with nucleophiles to the professor's emphasis on nucleophiles over electrophiles during class. Bhattacharyya and Harris' (2018) work suggests that the emphasis on identification of nucleophiles over electrophiles could also be attributed to the active voice typically used when describing the reaction step (i.e., saying "the nucleophile attacks the carbocation") based on the syntax of the electron pushing formalism. Respondents in our study explicitly noted that it was one of the lone electron pairs on ethanol (i.e., the nucleophile) that formed the bond with the carbocation (i.e., the electrophile), indicating they under-

المنارة للاستشارات
www.manaraa.com

stood something about the direction of electron flow; however, these respondents did not always label the nucleophile and electrophile with the proper chemical terminology.

Leaving group

Responses that mentioned the proton transfer step (e.g., "once the ethanol is attached to the t-butyl, another ethanol will remove the proton") without explanation were coded as Level 1. Many students described the proton-transfer step as occurring to "get rid of the charge" or "make the product neutral", which they related to stability of the product. This use of teleological reasoning was associated with Level 2 scores. An ideal Level 3 explanation for this step would describe the electronegative oxygen of a second ethanol molecule attracting the partially positive hydrogen, forming a new bond, with the electrons from the broken bond becoming a lone pair on oxygen. Though students have previously been able to describe acid–base proton-transfer reactions using the Lewis acid–base model in the context of a standard-curriculum organic chemistry course (Dood *et al.*, 2018) and causally in the context of a revised curriculum (Cooper *et al.*, 2016; Crandell *et al.*, 2018) when provided with a similar prompt, many students in this study simply omitted discussion about the acid–base proton-transfer step. It is likely that students were more focused on the first two steps of the reaction (i.e., the steps that occur in all S_N1 reactions) and either did not feel the need to explain the proton-transfer step or did not realize it was there.

Use of anthropomorphisms

Based on the literature, we expected a significant amount of anthropomorphism use in responses (Taber and Watts, 1996). Instances in our dataset ranged from humorous commentaries "Bromine is a very heavy and electronegative atom making it a perfect leaving group because it bares the negative charges very well and *honestly does better on its own living the single life*" (emphasis added), to more mundane "Bromine is a good leaving group because it *wants* to be on its own". Though studies have shown that

المنارة للاستشارات
www.manaraa.com

the use of anthropomorphisms in teaching can be harmful to developing understanding (Talanquer, 2013, 2007), anthropomorphism use can also be evidence for sound chemical reasoning as one could argue was demonstrated in the response that bromine is better "on its own living the single life". However, this is more the exception than the rule. Many students apply humanlike qualities to molecules and are unable to give reasoning for why a reaction occurs other than the molecule *wants* to. For example, when describing the final step (i.e., the proton-transfer step) of the reaction, many respondents stated that the compound did not *want* to have a charge, or the compound *wants* to be neutral. The word "want" occurred 223 times in our dataset, with very few instances of this anthropomorphism coupled with sound chemical reasoning.

Research has shown that use of anthropomorphisms during instruction can encourage student misunderstandings, as students take the human-like models literally (Talanquer, 2007, 2013) rather than appreciate the analogy intended. Results of our study echo this, as we found that many students wrote about atoms or molecules *wanting* without providing causal explanations alongside anthropomorphic explanations. Though experts understand that a molecule does not *want*, novices who are given these anthropomorphic explanations without instruction in how they relate to chemical reasoning may believe that such rationale is sufficient for understanding the chemical activity of the system.

Overall level assignments

Overall, there were a small (N = 113) number of responses scored as Level 1; this indicates respondents in our sample are able to articulate what is happening in the reaction mechanism and, to some degree, why a reaction is occurring at a molecular level (i.e., responses scored as Level 2 or Level 3). The prompt was designed to invoke explanations by asking what was happening and why it was happening separately (see Cooper *et al.*, 2016); thus, inclusion of an attempt to provide a rationale is expected given the structure of the assessment item. Most responses (N = 549, 53%) were coded as Level 2; 379 (36%) were coded as Level 3. This is expected, as students are heavily assessed on the what (i.e., draw the mechanism, predict

www.manaraa.com

the products) and less often asked why reactions occur. Caspari *et al.* (2018a) found that many students did not explain with high levels of complexity until after prompting by the interviewer; thus, it is possible that our students may be capable of inferring implicit features of the reaction and using those features to reason about the why, but simply did not feel it was necessary to include in their response because they had not previously been required to provide this type of explanation.

5.10.2 RQ2: Does lexical analysis lead to a logistic regression model for predicting level of explanation sophisitication in response to a constructed-response formative assessment item on explaining an S_N1 reaction mechanism?

Two binomial logistic regression models were used to predict the tri-level scores for each response. The first model (i.e., Model #1) codes Level 1 responses as *0* and Level 2 and 3 responses as *1*. The second model (i.e., Model #2) codes Level 1 and 2 responses as *0* and Level 3 responses as *1*. The results of the two models are combined to assign Level 1, 2, and 3 to responses. Following the iterative development process described in the Methods section, 91.6% overall accuracy was achieved with the training dataset; this includes an 85.0% accuracy for Level 1, a 90.3% accuracy for Level 2, and a 95.5% accuracy for Level 3. Incorrect assignments were within one level of the correct assignment; in other words, there were no Level 1 responses scored by the computer as Level 3 and there were no Level 3 responses scored by the computer as Level 1. When the same model was applied to the testing dataset, the overall prediction accuracy was 86.9%; the accuracy for Level 1 was 84.4%, the accuracy for Level 2 was 85.1%, and the accuracy for Level 3 was 90.6%. Again, all incorrect assignments were one level away from the correct assignment. A description of categories used in each model, including rules, types, words and phrases that comprise each category, is provided in Appendix 2. Regression coefficients (β), *p* values, and odds ratios for each category by model are reported in Table 5.4.

المنارة للاستشارات
www.manaraa.com

Table 5.4 Descriptive summary of categories by level classification for the training set and results of two binomial logistic regressions.

Category	Model #1					Model #2				
	L1 % (n=76)	L2/3 % (n=652)	β	OR	p	L1/2 % (n=456)	L3 % (n=272)	β	OR	p
Accept/donate electrons	22.4	73.0	-0.77	2.17^a		26.5	26.5	0.41	1.51	
Attraction	0	8.9	2.70	14.95		0	21.3	6.46	636.5	***
Absence of explanation	5.3	14.6	-2.92	18.51^a	***	11.8	16.5	-2.98	19.77^a	**
Bond breaks	1.3	31.8	1.42	4.12		0.7	75.4	0.52	1.69	
Bond electrons	90.1	7.8	0.11	1.12		25.9	0.7	0.29	1.34	
Bond forms	7.9	20.0	1.65	5.22		14.5	25.7	0.25	1.28	
Carbocation attacked	6.6	16.1	-0.19	1.20^a		12.9	18.8	0.53	1.69	
Carbocation	59.2	86.4	-0.10	1.11^a		81.4	87.1	-1.00	0.37	
Degree carbocation	34.2	52.6	1.06	2.89		48.7	54.0	-0.34	1.41^a	
Deprotonate	2.6	4.8	1.51	4.52		4.2	5.1	0.91	2.48	
Donate hydrogen	19.7	57.7	1.65	5.22		41.9	73.5	1.02	2.77	
"Don't know"	0	27.2	-7.41	1660^a	*	6.1	54.8	0.66	1.93	
Electron attack	1.3	34.4	-2.66	14.21^a	*	26.1	39.0	0.77	2.17	
Electron terminology	0	35.1	2.18	8.82	*	29.6	34.6	-0.09	1.09^a	
Electronegativity	36.8	44.0	-	-		47.6	36.0	4.22	68.24	***
Electrophile accepts	47.4	44.8	-.022	1.23^a		50.0	36.8	-0.64	1.91^a	
Eliminate charge	7.9	15.2	3.45	31.48	**	14.9	13.6	-0.52	1.69^a	
Good leaving group	0	49.9	-	-		41.5	50.0	-0.21	1.24^a	
Leaving group leaves	0	5.1	0.80	2.23		3.7	5.9	0.16	1.17	
Nucleophile attacks	1.3	2.3	-0.55	1.75^a		2.6	1.5	0.14	1.15	
Nucleophile/electrophile terminology	1.3	35.0	-0.14	1.15^a		27.4	38.2	-0.68	1.97â	
Opposites	9.2	1.1	-	-		2.4	1.1	4.83	125.8	***
Partial charges	4.0	24.1	-	-		17.5	29.4	7.64	2070	***
Positive/negative charges	0	9.7	3.95	51.73	**	10.8	5.1	-0.12	1.13^a	
Reaction terminology	0	9.4	0.23	1.26		8.6	8.1	-0.55	1.74^a	
Solvent terminology	10.5	21.3	0.33	1.39		14.5	29.8	-0.92	2.51^a	
Stability carbocation	2.6	10.4	0.68	1.97		7.0	14.0	0.06	1.06	
Stability terminology	0	6.0	-	-		0.4	13.6	-0.52	1.69^a	
Sterics	51.3	68.4	-	-		64.9	69.5	2.13	8.37	*
Temperature	30.3	30.2	1.95	7.04		31.1	28.7	0.60	1.82	
Wants	2.6	48.5	0.76	2.14		32.5	62.5	-0.35	1.42^a	
Weak/strong base	6.6	24.4	-	-		16.7	32.4	-0.06	1.06^a	
Weak/strong nucleophile	2.6	28.2	0.97	2.63		24.1	27.9	0.36	1.43	
constant			-0.65	0.52				-3.20	0.04	

[a]inverse odds ratio. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. For Model #1: $\chi^2(26) = 156.36$, p<0.0001.
For Model #2: $\chi^2(33) = 742.22$, p<0.0001.

Regression model 1

The first binomial logistic regression model differentiates Level *1* responses (coded as *0*) from Level 2 and Level 3 responses (coded as *1*). Significant predictors include *absence of explanation, "don't know", electron terminology, electron attack, eliminate charge,* and *positive/negative charges*. Unsurprisingly, the categories *absence of explanation* and '*'don't know"* were strong predictors for a response being scored as Level 1, with inverse odds ratios of 19 and 1,660, respectively. Thus, the odds of responses that included

المنارة للاستشارات
www.manaraa.com

the category *absence of explanation* being scored as Level 1 (as opposed to being scored as Level 2/Level 3) were 19 times higher than responses that did not include the category, and the odds of responses that claimed they did not know why the reaction is occurring (i.e., included the category *"don't know"*) being scored as Level 1 were 1,660 times higher with all other factors held constant. The category electron attack was also a strong predictor for a response being scored as Level 1, with an inverse odds ratio of 14. The nature of this category as a negative predictor is discussed further below. Use of electron terminology, talking about eliminating the charge in the reaction, and talking about positive and negative charges were positive predictors for a response being scored as Level 2 or Level 3. The odds of responses that included these categories being scored as Level 2 or Level 3 (as opposed to be begin scored as Level 1) were 9, 32, and 52 times higher, respectively, compared to responses that did not include these categories and holding all other factors constant.

Regression model 2

The second binomial logistic regression model differentiated Level 1 and 2 responses (coded as *0*) from Level 3 responses (coded as *1*). Significant positive predictors for Level 3 include *partial charges, attraction, electronegativity, sterics,* and *opposites*. *Absence of explanation* is a significant negative predictor with an inverse odds ratio of 20. The categories *partial charges, attraction,* and *electronegativity* describe reasoning that would be included in a Level 3 response. The odds of responses coded with these categories being scored are Level 3 (as opposed to being scored as Level 1/Level 2) are 2070, 637, and 68 times higher, respectively, compared to response not coded with these categories when holding all other factors constant. The odds of responses including the category *sterics* being scored as Level 3 (as opposed to being scored as Level 1/Level 2) were eight times higher than for responses that did not include *sterics*, holding all other variables constant. The category *opposites* includes a rule that looks for types *negative* and *positive* in any order, a rule that looks for the type *positive* and then the type *attack* (e.g., "the positive charge is

المنارة للاستشارات
www.manaraa.com

being attacked"), a rule that looks for the type *negative* and then the type *attack* (e.g., "the negative charge is attacking"), a rule that looks for the type opposite and then the type *charge*, and the word dipole. The odds of responses in this category being scored as Level 3 over Level 1/Level 2 were 126 times higher than responses that were not in this category when holding all other variables constant.

Overall level assignments

An 86.9% overall accuracy with the testing dataset suggests that we have developed a suitable set of predictive regression models for scoring the constructed-response item. While this level of accuracy exceeds reported accuracies in similar studies, we recommend that use of resultant scores be limited to formative assessments; this recommendation mirrors recommendations for use by others who have developed analogous computer-assisted scoring models (Moharreri *et al.*, 2014; Link *et al.*, 2017).

Discussion of electrons was important across all three levels. The use of electron terminology was a predictor for a response to be scored as Level 2 or Level 3. Typically, the use of electron terminology included explanation such as describing lone pairs attracting an electrophile (deeper-level explanation) or even bromine *wanting* to keep its electrons to itself (i.e., surface-level teleological explanation). The use of phrases where electrons *attack* was a negative predictor for Level 2 or 3, as the inclusion of a phrase indicating electrons are attacking is not enough to indicate an explanation of why the reaction is occurring. However, writing about electrons attacking may be a first step toward developing an explanation for why the attack is occurring. The categories *partial charges, attraction*, and *electronegativity* were also predictors for a Level 3 score.

Another important concept that differentiated a Level 1 score from Level 2 score is the use of charges in an explanation. The categories *eliminate charge* and *positive/negative charges* were significant predictors of Level 2 and Level 3 scores. *Partial charges* was a significant predictor of a Level 3 score. The category *opposites* was also a predictor of a Level 3 score; this is a category that refers to positive and negative

المنارة للاستشارات
www.manaraa.com

partial and formal charges interacting with each other. This corroborates findings from Anzovino and Bretz (2016) where charges were an important part of student understanding of nucleophiles and electrophiles. Most participants in the Anzovino and Bretz study, as well as Level 2 responses in our study, used charges to explain the reactivity of nucleophiles and electrophiles. Partial charges, which better explain reactivity than formal charges alone, were associated with Level 3 responses. Areas of high and low electron density (i.e., partial charges) interacting with each other indicate deeper reasoning (i.e., a Level 3 score), as partial charges are an implicit feature.

5.11 Implications

5.11.1 Implications for researchers

Formative assessment items can be developed to further evaluate students' explanations of scientific phenomena. Analysis of written assessment responses including through lexical analysis and logistic regression models can provide insight about student explanations. Though some time is required to develop logistic regression models, once models are developed, they can be used for formative assessment and to evaluate the impact of interventions on student responses to constructed-response items. Our analyses were conducted using Python, an open-source scripting language (seehttps://osf.io/muw3r/?view_only=7ceacdb889704083b098a4239235e2dd), making these tools more accessible. Researchers can develop predictive models without experiencing the financial barrier of purchasing software. In the field of organic chemistry, constructed-response items should be developed that cover a broader range of reaction types (e.g., S_N2, E1, E2, Diels-Alder). Analogous constructed-response items could allow researchers to more deeply understand student explanations of those reaction types and uncover persistent misunderstanding across a spectrum of reaction types. Future models could include more detailed analysis of student responses, such as level assignments for explanations of specific aspects of the reaction (e.g., leaving group, carbocation,

www.manaraa.com

nucleophile and electrophile, and proton transfer). Additionally, constructed-response items that include case comparison, such as those used by Bodé *et al.* (2019) and Caspari *et al.* (2018a), could be developed to further elicit student explanations and predictive logistic regression models could be developed for such items. Predictive models could be developed to score responses to prompts of the type used in this study based on the frameworks for levels of explanation used by Bodé *et al.* (2019) and Caspari *et al*. (2018a).

Resultant predictive models could be used to develop targeted interventions for students, such as the targeted intervention we developed to encourage students to explain what is happening in an acid−base proton-transfer mechanism and why using the Lewis acid−base model (Dood *et al.*, 2019). Instructional tools like our Lewis acid−base tutorial could be developed that remedy lack of understanding about specific topics. Such remedies should be grounded in implications of studies regarding those specific parts of the reaction mechanism. For example, in our study, we considered the four main parts of the reaction when evaluating the three Levels of explanation sophistication that have previously been explored by other researchers: leaving groups (Popova and Bretz, 2018), carbocations (Bodé *et al.*, 2019; Caspari *et al.*, 2018a), nucleophiles and electrophiles (Anzovino and Bretz, 2015, 2016), and acid−base proton transfer (Cartrette and Mayo, 2011; Cooper *et al.*, 2016; Crandell *et al.*, 2018). An understanding and synthesis of all of these topics is required for a complete understanding of the reaction mechanism (c.f., Cruz-Ramírez de Arellano and Towns, 2014). Our work could be used to develop learning tools that remedy lack of understanding of each of these parts individually and encourage a synthesis of the components to form a coherent causal rationale of reaction mechanisms. Application program interfaces (APIs) could be developed to analyze student responses in real time and provide learners with instructional tools based on categories that are missing from their responses and level of understanding. Students could respond to a prompt and have the response immediately scored by the API. Students would not see their score; the score, though, would be used to direct the program to provide students with a specific tutorial meant to assist them based on their current level of explanation sophistication. For example, if a student were scored at Level 1, a tutorial meant to

المنارة للاستشارات
www.manaraa.com

aid students in developing their explanation to a Level 2 explanation could be provided. Students scored at Level 2 initially could have a tutorial aimed at developing their current Level 2 explanation into a Level 3 explanation. Students already at Level 3 could receive a tutorial that takes them beyond the levels expressed in this model or a tutorial reinforcing the concepts they have already expressed in their explanation. There are many options for what a predictive model can code. The developer of the model can choose to code for specific text patterns, phrases, or terminology at their desired grain size. Models could be developed that identify the use of specific cause-and-effect relations or implicit properties. The work presented in this paper only scratches the surface of the possibilities for lexical analysis, logistic regression models, and APIs for use in organic chemistry courses.

5.11.2 Implications for instructors

Students should be assessed regarding molecular explanations about why reactions occur, as assessments provide students with a strong message about what is important in the course (Holme *et al.*, 2010). Other researchers have called for instructors to use assessments that ask students to engage in reasoning about the why (Caspari *et al.*, 2018a; Cooper, 2015; Cooper *et al.*, 2016). If an instructor wishes to summatively assess students on their explanations of phenomena, it is important to change their teaching strategies to aid development of such skills in a low-stakes environment. One way to develop such skills in a formative way is through the use of assignments which use logistic regression techniques to provide large numbers of students with immediate feedback. The model presented in this study is available for use by instructors in their courses (https://osf.io/muw3r/?view_only=7ceacdb889704083b098a4239235e2dd). Given the non-perfect accuracy of our scoring tool, we recommend, as have others, that our predictive models be used to score the constructed-response items only in formative assessment contexts. Model output includes an overall Level assignment as well as category codes; instructors can use this depth of information to tailor lectures to address areas missing from many students' explanations. For example, if the majority of student

المنارة للاستشارات
www.manaraa.com

responses are assigned Level 1, it would be important to help develop students' explanations in such a way that students will begin to explain why the reaction is occurring rather than just describing the reaction. The absence of explanation category, a negative predictor for a Level 2 or Level 3 score, includes responses that lack the types associated with explanations of why the reaction is occurring. Such explanations could include types such as attract, electronegativity, partial, stability, degree, strong, and weak. A professor can use information about their students' responses and emphasize these topics in their lecture.

In addition, students could receive a list of categories coded for their response as well as ideal categories that were not coded by an API immediately after completing a constructed-response item. This fine-grain level of information coupled with targeted-instructional materials assigned to the student by the API based on their written response could be used by the student to better understand what level of explanation is expected of them and to direct their attention to areas they have missed.

Preparing students for success in a curriculum where examinations require them to explain the why rather than simply regurgitating memorized information may require additional instruction on learning how to learn and study. For example, one way to encourage and teach self-regulated learning (Zimmerman and Martinez-Pons, 1988) is through a intervention module such as the Growth and Goals project from the University of Ottawa (Flynn, 2016).

Our data suggest that students continue to use anthropomorphisms in their explanations. Instructors should limit use of anthropomorphisms and teleology when describing chemical phenomena, encouraging students to develop sound chemical reasoning. Deeper-level chemical explanations including the use of explicit features to infer implicit features should be encouraged. Talanquer (2013) suggests specifically discussing the differences between teleological and more chemically sound reasoning, such as causal mechanistic reasoning, for the same system during class, so that students learn to develop explanations and apply reasoning in a sound chemical manner.

www.manaraa.com

5.12 Limitations

There are some key limitations of our study: homogeneity of the sample, lack of ability to ask follow-up questions to prompt deeper explanations, level of accuracy of the scoring model, and focus on one specific reaction type. The response data were collected from students who took organic chemistry with one of three instructors during two different semesters; these students experienced one curriculum setting at one institution. Different curricula have been reported in the literature for the postsecondary yearlong course in organic chemistry (e.g., the Mechanisms before Reactions curriculum; Flynn and Ogilvie, 2015); students in other contexts may respond differently to the constructed-response item and may require modification of the coding scheme based on different emphases and levels of explanation sophistication. Despite the homogeneity of our responses, our model is generalizable as our findings mirror those of more interview-based studies and human-scored assessments reported in the literature.

Additionally, as the constructed-response item was administered using a survey, there was not an opportunity to ask students follow-up questions that may have prompted the use of deeper explanations (i.e, Level 3). Caspari *et al.* (2018a) were able to observe high levels of complexity in student explanations, but this complexity typically did not occur until after follow-up questions were asked by the interviewer. Given this finding, it is possible we could have elicited more Level 3 responses from students had the study used interviews.

Another limitation is the overall accuracy of the predictive model: 86.9%. Our accuracy level is similar to other published predictive models (Dood *et al.*, 2018; Haudek *et al.*, 2012; Prevost *et al.*, 2016, 2012) developed for similar purposes, but is still non-perfect. We thus reiterate our recommendation that the computer-assisted scoring be used for formative assessment purposes.

We report a scoring model for one specific reaction type (i.e., S_N1) with minimally varied starting materials and reagents. As students learn many reactions and several mechanism types in organic chemistry,

المنارة للاستشارات
www.manaraa.com

it would be useful to expand the tool to include more S_N1 reactions and develop additional tools for use with other reaction types (e.g., S_N2, E1, E2). Items could also be developed that ask students to make comparisons about the feasibility of different reaction mechanisms (Bodé *et al.*, 2019; Caspari *et al.*, 2018a). This may require the construction of multiple models, as the same coding scheme may not fit with all mechanism types. For example, the reaction studied here includes an acid−base step at the end. Not all substitution and elimination reactions require this step. A reaction mechanism of the same type (i.e., S_N1) that does not include this step would likely not fit the current predictive model as is, but the model could be modified to accommodate this difference.

5.13 Conclusion

This study has shown that three levels of explanation emerged for our sample of students in response to a constructed-response item on the S_N1 reaction mechanism. The assessment item and corresponding logistic regression models we presented in this study revealed themes in student understanding that corroborated those found by others in the literature. Results of our study can be used to modify teaching practices and develop instructional tools to help students develop a deeper understanding of S_N1 reaction mechanisms.

5.14 Conflicts of interest

There are no conflicts to declare.

5.15 References

Abrams, E., Southerland, S., (2001), The how's and why's of biological change: How learners neglect physical mechanisms in their search for meaning. *Int. J. Sci. Educ.*, **23**, 1271–1281.

المنارة للاستشارات
www.manaraa.com

Anzovino, M.E., Bretz, S.L., (2016), Organic chemistry students' fragmented ideas about the structure and function of nucleophiles and electrophiles: a concept map analysis. *Chem. Educ. Res. Pract.*, **17**, 1019–1029.

Anzovino, M.E., Bretz, S.L., (2015), Organic chemistry students' ideas about nucleophiles and electrophiles: the role of charges and mechanisms. *Chem. Educ. Res. Pract.*, **16**, 797–810.

Becker, N., Noyes, K., Cooper, M., (2016), Characterizing Students' Mechanistic Reasoning about London Dispersion Forces. *J. Chem. Educ.*, **93**, 1713–1724.

Bhattacharyya, G., (2013), From source to sink: mechanistic reasoning using the electron-pushing formalism. *J. Chem. Educ.*, **90**, 1282–1289.

Bhattacharyya, G., Bodner, G.M., (2005), "It gets me to the product": how students propose organic mechanisms. *J. Chem. Educ.*, **82**, 1402–1407.

Bhattacharyya, G., Harris, M.S., (2018), Compromised structures: Verbal descriptions of mechanism diagrams. *J. Chem. Educ.*, **95**, 366–375.

Bodé, N.E., Deng, J.M., Flynn, A.B., (2019), Getting past the rules and to the WHY: Causal mechanistic arguments when judging the plausibility of organic reaction mechanisms. *J. Chem. Educ.*, **96**, 1068–1082.

Cartrette, D.P., Mayo, P.M., (2011), Students' understanding of acids/bases in organic chemistry contexts. *Chem. Educ. Res. Pract.*, **12**, 29–39.

Caspari, I., Kranz, D., Graulich, N., (2018a), Resolving the complexity of organic chemistry students' reasoning through the lens of a mechanistic framework. *Chem. Educ. Res. Pract.*, **19**, 1117–1141.

Caspari, I., Weinrich, M.L., Sevian, H., Graulich, N., (2018b), This mechanistic step is "productive": organic chemistry students' backward-oriented reasoning. *Chem. Educ. Res. Pract.*, **19**, 42–59.

Cooper, M., (2015), Why Ask Why? *J. Chem. Educ.*, **92**, 1273–1279.

Cooper, M., Kouyoumdjian, H., Underwood, S., (2016), Investigating students' reasoning about acid–base reactions. *J. Chem. Educ.*, **93**, 1703–1712.

Crandell, O., Kouyoumdjian, H., Underwood, S., Cooper, M., (2018), Reasoning about reactions in organic chemistry: Starting It in general chemistry. *J. Chem. Educ.*, **96**, 213-226.

Cruz-Ramírez de Arellano, D., Towns, M.H., (2014), Students' understanding of alkyl halide reactions in undergraduate organic chemistry. *Chem. Educ. Res. Pract.*, **15**, 501–515.

Dood, A.J., Fields, K.B., Cruz-Ramírez de Arellano, D., Raker, J.R., (2019), Development and evaluation of a Lewis acid–base tutorial for use in postsecondary organic chemistry courses. *Can. J. Chem.*, **97**, 711-721.

Dood, A.J., Fields, K.B., Raker, J.R., (2018), Using lexical analysis to predict Lewis acid–base model use in responses to an acid–base proton-transfer reaction. *J. Chem. Educ.*, **95**, 1267–1275.

Flynn, A., (2016), Growth and Goals Project Details [WWW Document]. Flynn Res. Group,.

Flynn, A.B., Ogilvie, W.W., (2015), Mechanisms before reactions: A mechanistic approach to the organic chemistry curriculum based on patterns of electron flow. J. *Chem. Educ.*, **92**, 803–810.

Glaser, B.G., Strauss, A.L., (1967), The discovery of grounded theory: strategies for qualitative research. Aldine Publishing, Chicago.

Grove, N.P., Cooper, M.M., Rush, K.M., (2012), Decorating with arrows: Toward the development of representational competence in organic chemistry. *J. Chem. Educ.*, **89**, 844–849.

Haudek, K.C., Prevost, L.B., Moscarella, R.A., Merrill, J., Urban-Lurain, M., (2012), What are they thinking? Automated analysis of student writing about acid−base chemistry in introductory biology. *CBE Life Sci. Educ.*, **11**, 283–293.

Holme, T., Bretz, S.L., Cooper, M., Lewis, J., Paek, P., Pienta, N., Stacy, A., Stevens, R., Towns, M., (2010), Enhancing the role of assessment in curriculum reform in chemistry. *Chem. Educ. Res. Pract.*, **11**, 92–97.

www.manaraa.com

Kaplan, J.J., Haudek, K.C., Ha, M., Rogness, N., Fisher, D.G., (2014), Using lexical analysis software to assess student writing in statistics. *Technol. Innov. Stat. Educ.*, **8**.

Kermack, W.O., Robinson, R., (1922), LI.—An explanation of the property of induced polarity of atoms and an interpretation of the theory of partial valencies on an electronic basis. *J. Chem. Soc. Trans.*, **121**, 427–440.

Koslowski, B., (1996), Theory and Evidence the Development of Scientific Reasoning.

Lemke, J.L., (1990), Talking Science: Language, Learning, and Values. Ablex Publishing Corporation, 355 Chestnut Street, Norwood, NJ 07648 (hardback: ISBN-0-89391-565-3; paperback: ISBN-0-89391-566-1).

Link, S., Chukharev-Hudilainen, E., Ranalli, J., (2017), Automated writing evaluation for formative assessment of second language writing: investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educ. Psychol.*, **37**, 8–25.

Moharreri, K., Ha, M., Nehm, R.H., (2014), EvoGrader: an online formative assessment tool for automatically evaluating written evolutionary explanations. *Evol. Educ. Outreach*, **7**, 15.

Moreira, P., Marzabal, A., Talanquer, V., (2019), Using a mechanistic framework to characterise chemistry students' reasoning in written explanations. *Chem. Educ. Res. Pract.*, **20**, 120–131.

National Research Council, (2011), A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., (2011), Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, **12**, 2825-2830.

Popova, M., Bretz, S.L., (2018), Organic chemistry students' understandings of what makes a good leaving group. *J. Chem. Educ.*, **95**, 1094–1101.

Prevost, L.B., Haudek, K., Urban-Lurain, M., Merrill, J., (2012), Examining student constructed explanations of thermodynamics using lexical analysis, in: 2012 Frontiers in Education Conference Proceedings. Presented at the 2012 Frontiers in Education Conference Proceedings, pp. 1–6.

Prevost, L.B., Haudek, K.C., Norton Henry, E., Urban-Lurain, M., Berry, M.C., (2013), Automated text analysis facilitates using written formative assessments for just-in-time teaching in large enrollment courses. Presented at the 120th ASEE Annual Conference & Exposition.

Prevost, L.B., Smith, M.K., Knight, J.K., (2016), Using student writing and lexical analysis to reveal student thinking about the role of stop codons in the central dogma. *CBE-Life Sci. Educ.*, **15**, ar65: 1-13.

Russ, R.S., Scherr, R.E., Hammer, D., Mikeska, J., (2008), Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Sci. Educ.*, **92**, 499–525.

Schauble, L., (1996), The development of scientific reasoning in knowledge-rich contexts. *Dev. Psychol.*, **32**, 102–119.

Sevian, H., Talanquer, V., (2014), Rethinking chemistry: a learning progression on chemical thinking. *Chem. Educ. Res. Pract.*, **15**, 10–23.

Shen, J., Liu, O.L., Sung, S., (2014), Designing interdisciplinary assessments in sciences for college students: An example on osmosis. *Int. J. Sci. Educ.*, **36**, 1773–1793.

Sperber, D., Premack, D., Premack, A.J., (1996), Causal Cognition: A Multidisciplinary Debate. Oxford University Press.

Strauss, A.L., Corbin, J.M., (1990), Basics of qualitative research: grounded theory procedures and techniques. Sage Publications, Newbury Park, Calif.

Strickland, A.M., Kraft, A., Bhattacharyya, G., (2010), What happens when representations fail to represent? Graduate students' mental models of organic chemistry diagrams. *Chem. Educ. Res. Pract.*, **11**, 293–301.

المنارة للاستشارات
www.manaraa.com

Taber, K.S., Watts, M., (1996), The secret life of the chemical bond: students' anthropomorphic and animistic references to bonding. *Int. J. Sci. Educ.*, **18**, 557–568.

Talanquer, V., (2018), Importance of understanding fundamental chemical mechanisms. *J. Chem. Educ.*, **95**, 1905–1911.

Talanquer, V., (2013), When atoms want. *J. Chem. Educ.*, **90**, 1419–1424.

Talanquer, V., (2007), Explanations and teleology in chemistry education. *Int. J. Sci. Educ.*, **29**, 853–870.

Tamir, P., Zohar, A., (1991), Anthropomorphism and teleology in reasoning about biological phenomena. *Sci. Educ.*, **75**, 57–67.

Wright, L., (1976), Teleological Explanations: An Etiological Analysis of Goals and Functions. University of California Press.

Wright, L., (1972), Explanation and teleology. *Philos. Sci.*, **39**, 204–218.

Yan, F., Talanquer, V., (2015), Students' ideas about how and why chemical reactions happen: Mapping the conceptual landscape. *Int. J. Sci. Educ.*, **37**, 3066–3092.

Zimmerman, B.J., Martinez-Pons, M., (1988), Construct validation of a strategy model of student self-regulated learning. *J. Educ. Psychol.*, **80**, 284–90.

Zimmerman, C., (2000), The development of scientific reasoning skills. *Dev. Rev.*, **20**, 99–149.

www.manaraa.com

Chapter 6

Conclusion

The work presented in this dissertation seeks to more deeply understand and evaluate student descriptions and explanations of organic chemistry reaction mechanisms. Beginning with a constructed-response item developed by Cooper *et al.* (2016) which asked what is happening in an acid−base proton transfer reaction and why, my work expanded upon that work to include automated scoring, a targeted learning tool, and an additional reaction type (i.e., S_N1). In Chapter 3, the work provides evidence that use of the Lewis acid−base model may increase students' success in an organic chemistry course. The work discussed in Chapters 3 and 5 demonstrates that logistic regression models can be used to predict scores for select written assignments in chemistry courses (i.e., predict Lewis acid−base usage or predict level of explanation sophistication). Text analysis can also be used to develop and evaluate targeted interventions, which was demonstrated in Chapter 4. The Lewis acid−base tutorial is only one example of many possibilities of what can be done for education and education research using automated text analysis. In this chapter, I will summarize my work and describe in detail the broader implications of the work for researchers and instructors.

6.1 Summary of results

In Chapter 3, I described a constructed-response item which asked what is happening and why in an acid−base proton transfer reaction mechanism. Student responses were coded for use or non-use of the

www.manaraa.com

Lewis acid–base model. I found that when comparing whether or not students used the Lewis acid–base model in their response to their exam scores on an exam with acid–base content, students who used the Lewis acid–base model scored significantly higher on acid–base related items and on the exam overall. These findings indicate that using the Lewis acid–base model when responding to this constructed-response item may be important for student success in an organic chemistry class. In order to make the item practical to administer to a large class, I developed a logistic regression model that predicts Lewis acid–base usage with high accuracy through automated text analysis.

To further the utility of the Lewis acid–base constructed-response item, I developed a tutorial to facilitate the construction of student knowledge of the Lewis acid–base model based on what is known about students' understanding of acids and bases in the literature (Chapter 4). Cloned versions of the constructed-response item developed in Chapter 3 were used before and after administering the tutorial to evaluate the impact of the tutorial on students' use of the Lewis acid–base model. The tutorial included units covering: (1) the definition of Lewis acids and bases (Cartrette and Mayo, 2011), (2) the amphoteric property of water (Drechsler and Van Driel, 2008; Garnett *et al.*, 1995), (3) acid–base reactions without a proton transfer (Tarhan and Sesen, 2012), and (4) integration of concepts (Taagepera and Noori, 2000). Rather than using a control group who did not receive the tutorial, a McNemar test (McNemar, 1947) was used to evaluate the pre/post design. The tutorial was administered to three samples of sophomore organic chemistry students: first semester before summative assessment on acids and bases, first semester after summative assessment on acids and bases, and second semester students. A positive impact on use of the Lewis acid–base model after the tutorial was observed for all populations. Additionally, the first semester students who originally completed the tutorial before summative assessment on acids and bases were given a third iteration of the constructed-response item after a three-week delay. The difference in use of the Lewis acid–base model in these responses compared to the responses immediately following the tutorial was not significant, suggesting that the tutorial may have had a lasting impact.

المنارة للاستشارات
www.manaraa.com

A second constructed-response item and predictive logistic regression model were presented in Chapter 5. This time, the item asked what is happening and why in an S_N1 reaction. The item was coded for three levels of explanation sophistication. Level 1 included responses that described what was happening in the reaction but not any explanation of why the reaction occurs. Level 2 responses included a surface level explanation using only explicit features of the reaction or using implicit features with no deeper explanation, which might indicate a memorized pattern. Level 3 included a deeper explanation using implicit features inferred from explicit features of the reaction (e.g., discussing topics like electron density and partial charges as reasons for why the reaction occurs). There were four main parts of the reaction that students discussed in their responses: leaving group, nucleophile and electrophile, carbocation, and proton transfer. I developed two binomial logistic regression models which were combined to predict the level of explanation sophistication for responses with high accuracy. Chapter 5 also reports the development of a Python program which uses four algorithms to develop categories which are used as predictors for a logistic regression model: equivalents, types, rules, and categories. This program is freely available online (https://osf.io/muw3r/) and can be used to develop predictive models for other constructed-response items.

6.2 Implications for researchers

Written assessments can reveal students' understanding or lack thereof.

The use of constructed-response items in this dissertation revealed much information about how students describe and explain acid–base proton-transfer and S_N1 reaction mechanisms. The most telling finding is that many students were unable to explain *why* the reactions occur at a deep level. Researchers can use written assessments similar to those presented in this dissertation for other reaction mechanism types (e.g., S_N2, E1, E2, pericyclic, rearrangement, radical) to further understand how students explain the mechanisms

www.manaraa.com

of different reaction types and the specific parts of each mechanism. Similar written assessments can be used to ask students to explain any number of scientific phenomena.

Constructed-response items that go beyond simply describing and explaining a mechanism could also be developed. Graulich and Schween (2018) described the idea of items with contrasting cases in organic chemistry. Contrasting cases items take two similar cases and ask students to compare/contrast some aspect of the cases. For example, one could ask which carbocation is favored and why but include two tertiary carbocations with other differences besides degree (Caspari *et al.*, 2018a). This encourages students to think more deeply when searching for an explanation behind carbocation stability by going beyond degree of the carbocation. Items that present students with two similar mechanistic steps and ask them to describe which is more likely to occur and why could also be used (e.g., Bodé *et al.*, 2019) to encourage students to develop scientific argumentation skills. Researchers can use items that have contrasting cases or comparisons to more deeply probe student understanding of specific parts of mechanisms. The possibilities for question types and topics for constructed-response items are endless.

Predictive text analysis models can be developed and used to evaluate large data sets of students' written responses.

For the constructed-response items in this dissertation, predictive models were developed to score responses based on coding schemes that arose from the data. Once predictive models are developed, they can be used to efficiently score new data sets of responses to the same item. These constructed-response items can then be used to track students' progress over time or to compare the scores of students' responses to other measures (e.g., exam scores, ability to solve certain problems) in very little time.

المنارة للاستشارات
www.manaraa.com

Automated text analysis can be used to evaluate the effectiveness of research-based interventions.

After developing constructed-response items and determining areas of student difficulty, targeted interventions can be developed using the information gathered from student data and from the research literature. Like I demonstrated in Chapter 4, predictive models can be used to evaluate the effectiveness of interventions. After the contents of the dataset provided a clear scoring scheme (i.e., use or non-use of the Lewis acid−base model), a targeted intervention was developed to help students construct an explanation which would be scored at a higher level (in this case, use of the Lewis model). The predictive model was used to evaluate students' use of the Lewis model before and after the intervention. Use of the predictive model was an efficient way to evaluate the intervention because hundreds responses could be input into the predictive model and scored within seconds. Researchers can use the same process of developing a predictive model and intervention. They can then use the predictive model to evaluate the effectiveness of the intervention.

6.3 Implications for instructors

Instructors must emphasize the importance of using the Lewis acid−base model when explaining acid−base reaction mechanisms.

It has been suggested that the Lewis model is the most useful acid−base model for students to use in organic chemistry (e.g., Bhattacharyya, 2013; Nataro *et al.*, 2004; Shaffer, 2006), and Cooper *et al.* (2016) found that students who used the Lewis model were more likely to draw correct mechanistic arrows. Adding evidence to this claim, the work in Chapter 3 indicated that use of the Lewis acid−base model when explaining the mechanism of an acid−base proton-transfer reaction was associated with higher scores on an organic chemistry examination. Given the growing evidence that learning to apply the Lewis model may correlate with better performance on organic chemistry tasks, instructors should emphasize the Lewis model

www.manaraa.com

when explaining acid−base reactions in organic chemistry. Researchers have also encouraged emphasizing electronic effects when explaining reactions (Caspari *et al.*, 2018a) and explaining acid−base reactions at an electronic level.

Instructors should be clear that atoms do not want *and that saying something* wants *to occur is not a valid reason for a reaction occurring.*

In responses to the acid−base proton-transfer and S_N1 reaction mechanisms prompts, students described the *wants* and *needs* of atoms when explaining why the reaction occurs. This is an example of anthropomorphic reasoning (i.e., attributing human qualities to non-human entities). Talanquer (2013) argued that the use of such reasoning can give people a false sense of understanding so that they do not seek out deeper understanding. Experts are aware that attributing *wants* and *needs* to atoms is a metaphor, but my data suggest that students are not always aware that atoms do not *want*. Many students used the wants and needs of atoms as a way to explain why reactions occurred. Instructors should clearly explain that atoms do not want and that there are deeper, causal explanations of why reactions occur.

Short writing assignments coupled with targeted interventions can positively impact students' explanations of organic chemistry reaction mechanisms.

In Chapter 4, a targeted intervention to increase use of the Lewis acid−base model in responses to a constructed-response item positively impacted use of the Lewis model in students' responses. This tutorial can be used in the conjunction with instruction on acids and bases to help students construct and apply knowledge about Lewis acids and bases. This tutorial sets a precedent for other tutorials which can be developed through similar methods and used in the classroom. These tutorials can help instructors include a written explanation of mechanisms portion to the curriculum along with targeted interventions to improve students' explanations of organic chemistry reaction mechanisms.

المنارة للاستشارات
www.manaraa.com

6.4 Overall implications

Science education, as a whole, must consider the use of writing as a means of assessment if causal reasoning is to be emphasized.

The *Framework for K-12 Science Education* names eight practices of scientists that should be taught in science classrooms, including constructing explanations and engaging in scientific argumentation. Engaging students with writing in the science classroom is vital to engaging students with such science practices. Becker *et al.* (2016) are among the growing list of researchers (e.g., Becker *et al.*, 2016; Bodé *et al.*, 2019; Caspari *et al.*, 2018a, 2018b; Cooper, 2015; Cooper *et al.*, 2016; Crandell *et al.*, 2018) who have called for engaging students in the construction of scientific explanations, stating that it “affords them an opportunity to engage (in scaffolded form) in the practices of science, that is, the activities that scientists engage in as they investigate the world” (p. 1714). It has become clear that simply asking students to solve problems in science courses is not enough for them to develop causal reasoning (e.g., for organic chemistry mechanisms; Bhattacharyya and Bodner, 2005; Ferguson and Bodner, 2008; Grove *et al.*, 2012). Thus, requiring students to provide written explanations can provide insight into students’ understanding of causality of phenomena and assert to them the importance of understanding causality.

Larger scale writing to learn (WTL) assignments can encourage learning in the sciences (Bangert-Drowns *et al.*, 2004; Connolly and Vilardi, 1989; Keys, 1999; Prain, 2006; Reynolds *et al.*, 2012). WTL assignments have been used successfully in chemistry courses (e.g., Finkenstaedt-Quinn *et al.*, 2017; Moon *et al.*, 2018, 2019; Shultz and Gere, 2015) including organic chemistry courses (Nicotera *et al.*, 2001; Schmidt-McCormack *et al.*, 2019; Wilson, 1994). Instructors should consider constructed-response items and WTL assignments in order to provide students with the opportunity to engage with the scientific practices of constructing explanations and argumentation from evidence as formative and summative assessments.

المنارة للاستشارات
www.manaraa.com

Lexical analysis and predictive models have the potential to change the nature of assigning constructed-response items in the classroom.

Time required to provide feedback is a barrier for the implementation of explanation-type assessments in large lecture courses; the time required for instructors to provide feedback for a large number of written assignments is great and creates a delay in receiving feedback for students. The tools presented in this dissertation can be used by instructors in their courses to provide students with immediate feedback on written responses (i.e., category information and overall level of explanation or Lewis acid−base model usage) and provide instructors with immediate feedback on the overall responses of their class so instructors can tailor instruction based on the results. For example, if students are administered the acid−base proton transfer item and only one-third of students are coded as using the Lewis acid−base model, the instructor should decide to reiterate their instruction on the Lewis acid−base model and how to use it in the context of explaining reaction mechanisms. The immediate feedback can also be helpful for students. Students can look at certain categories they have omitted or use their level information to guide their studying so they can reach the expected explanation depth. Additionally, the tutorial can be administered in organic chemistry classes to help students construct explanation of acid−base reactions using the Lewis model.

The ability to develop and use predictive models to score constructed-response items is vital for the use of constructed-response items in science education to become standard. Though the utility of writing in science course has been exemplified through the use of short constructed-response items and WTL assignments, such assignments can have diminished utility in courses with high enrollment where instructor may be unable to provide students with high quality, timely feedback. Predictive logistic regression models can address this issue and give instructors and students immediate feedback on student writing. Predictive models can allow for the use of writing as a learning tool in courses where it was previously infeasible and allow for an increased use of writing in courses where some writing is currently used.

المنارة للاستشارات
www.manaraa.com

The future of science education should include open-access instructional materials which use writing, automated text analysis, and research-based targeted interventions for students.

The Automated Analysis of Constructed Response (AACR) group at Michigan State University (beyondmultiplechoice.org) provides instructors with access to a database of constructed-response items that can be used in high school or undergraduate classrooms, though the items are mainly focused on biology concepts. These items can be automatically scored using the website and instructors are provided with a report for their students' responses. Similar to this database, I envision a large database containing thousands of constructed-response items developed by researchers that are part of an adaptive system which meets students at their current level based on their written responses. The system would provide students with targeted resources to help them achieve a deep understanding of concepts and increase the sophistication of their response.. All interventions would be based in the current literature about students' understanding of the concepts being taught and students would be able to work through the items at their own pace. This system would allow students to participate in the scientific practices of constructing explanations and engaging in argument from evidence without the constraint of delayed feedback.

6.5 Summary

The work presented in this dissertation paves the way for the use of computerized scoring of written responses in organic chemistry courses. The study in Chapter 3 takes a previously developed constructed-response item (Cooper *et al.*, 2016) and develops a predictive model to determine whether a student's response is using the Lewis acid−base model or not. The study in Chapter 4 uses the predictive model reported in Chapter 3 to guide the development and evaluation of an intervention designed to encourage use of the Lewis acid−base model in explanations of acid−base proton-transfer reaction mechanisms. This sets a precedent for how predictive models can be used as a tool to evaluate the effectiveness of literature-based

المنارة للاستشارات
www.manaraa.com

interventions. The study in Chapter 5 develops a predictive model for an S_N1 reaction mechanism item asking what happens in the reaction and why. This demonstrates that predictive logistic regression models can be developed for multiple reaction types and can be used to code for different ideas, as the new model codes for three levels of explanation sophistication as opposed to just use or non-use of the Lewis model.. I hope that future researchers can use my work as a precedent to develop further predictive models for other reaction types and question types and to develop and evaluate targeted learning tools to be used when teaching organic chemistry.

6.6 References

Bangert-Drowns, R.L., Hurley, M.M., Wilkinson, B., (2004), The effects of school-based Writing-to-Learn interventions on academic achievement: A meta-analysis. *Rev. Educ. Res.*, **74**, 29–58.

Becker, N., Noyes, K., Cooper, M., (2016), Characterizing students' mechanistic reasoning about London Dispersion Forces. *J. Chem. Educ.*, **93**, 1713–1724.

Bhattacharyya, G., (2013), From source to sink: mechanistic reasoning using the electron-pushing formalism. *J. Chem. Educ.*, **90**, 1282–1289.

Bhattacharyya, G., Bodner, G.M., (2005), "It gets me to the product": how students propose organic mechanisms. *J. Chem. Educ.*, **82**, 1402–1407.

Bodé, N.E., Deng, J.M., Flynn, A.B., (2019), Getting past the rules and to the WHY: Causal mechanistic arguments when judging the plausibility of organic reaction mechanisms. *J. Chem. Educ.*, **96**, 1068–1082.

Cartrette, D.P., Mayo, P.M., (2011), Students' understanding of acids/bases in organic chemistry contexts. *Chem. Educ. Res. Pract.*, **12**, 29–39.

Caspari, I., Kranz, D., Graulich, N., (2018a), Resolving the complexity of organic chemistry students' reasoning through the lens of a mechanistic framework. *Chem. Educ. Res. Pract.*, **19**, 1117–1141.

Caspari, I., Weinrich, M.L., Sevian, H., Graulich, N., (2018b), This mechanistic step is "productive": organic chemistry students' backward-oriented reasoning. *Chem. Educ. Res. Pract.*, **19**, 42–59.

Connolly, P., Vilardi, T., (1989), Writing To Learn Mathematics and Science. Teachers College Press, 1234 Amsterdam Ave.

Cooper, M., (2015), Why Ask Why? *J. Chem. Educ.*, **92**, 1273–1279.

Cooper, M., Kouyoumdjian, H., Underwood, S., (2016), Investigating students' reasoning about acid–base reactions. *J. Chem. Educ.*, **93**, 1703–1712.

Crandell, O., Kouyoumdjian, H., Underwood, S., Cooper, M., (2018), Reasoning about reactions in organic chemistry: Starting it in general chemistry. *J. Chem. Educ.*, **96**, 213–226.

Drechsler, M., Van Driel, J., (2008), Experienced teachers' pedagogical content knowledge of teaching acid–base chemistry. *Res. Sci. Educ.*, **38**, 611–631.

Ferguson, R., Bodner, G.M., (2008), Making sense of the arrow-pushing formalism among chemistry majors enrolled in organic chemistry. *Chem. Educ. Res. Pract.*, **9**, 102–113.

المنارة للاستشارات
www.manaraa.com

Finkenstaedt-Quinn, S.A., Halim, A.S., Chambers, T.G., Moon, A., Goldman, R.S., Gere, A.R., Shultz, G.V., (2017), Investigation of the influence of a Writing-to-Learn assignment on student understanding of polymer properties. *J. Chem. Educ.*, **94**, 1610–1617.

Garnett, Patrick J., Garnett, Pamela J., Hackling, M.W., (1995), Students' alternative conceptions in chemistry: A review of research and implications for teaching and learning. *Stud. Sci. Educ.*, **25**, 69–96.

Graulich, N., Schween, M., (2018), Concept-oriented task design: Making purposeful case comparisons in organic chemistry. *J. Chem. Educ.*, **95**, 376–383.

Grove, N., Cooper, M., Rush, K., (2012), Decorating with arrows: Toward the development of representational competence in organic chemistry. *J. Chem. Educ.*, **89**, 844–849. Keys, C.W., (1999), Revitalizing instruction in scientific genres: Connecting knowledge production with writing to learn in science. *Sci. Educ.*, **83**, 115–130.

McNemar, Q., (1947), Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, **12**, 153–157.

Moon, A., Moeller, R., Gere, A.R., Shultz, G.V., (2019), Application and testing of a framework for characterizing the quality of scientific reasoning in chemistry students' writing on ocean acidification. *Chem. Educ. Res. Pract.*, **20**, 484-494.

Moon, A., Zotos, E., Finkenstaedt-Quinn, S., Gere, A.R., Shultz, G., (2018), Investigation of the role of writing-to-learn in promoting student understanding of light–matter interactions. *Chem. Educ. Res. Pract.*, **94**, 1610-1617.

Nataro, C., Ferguson, M.A., Bocage, K.M., Hess, B.J., Ross, V.J., Swarr, D.T., (2004), Lewis acid−base, molecular modeling, and isotopic labeling in a sophomore inorganic chemistry laboratory. *J. Chem. Educ.*, **81**, 722–724.

Nicotera, C.L., Shibley, I.A., Milakofsky, L.K., (2001), Incorporating a substantial writing assignment into organic chemistry: Library research, peer review, and assessment. *J. Chem. Educ.*, **78**, 50-53.

Prain, V., (2006), Learning from writing in secondary science: Some theoretical and practical implications. *Int. J. Sci. Educ.*, **28**, 179–201.

Reynolds, J.A., Thaiss, C., Katkin, W., Thompson, R.J., Wright, R.L., (2012), Writing-to-Learn in undergraduate science education: A community-based, conceptually driven approach. *CBE-Life Sci. Educ.*, **11**, 17–25.

Schmidt-McCormack, J.A., Judge, J.A., Spahr, K., Yang, E., Pugh, R., Karlin, A., Sattar, A., Thompson, B.C., Gere, A.R., Shultz, G.V., (2019), Analysis of the role of a writing-to-learn assignment in student understanding of organic acid–base concepts. *Chem. Educ. Res. Pract.*, **20**, 383-398.

Shaffer, A.A., (2006), Let us give Lewis acid−base theory the priority it deserves. *J. Chem. Educ.*, **83**, 1746–1749.

Shultz, G.V., Gere, A.R., (2015), Writing-to-Learn the nature of science in the context of the Lewis Dot Structure model. *J. Chem. Educ.*, **92**, 1325–1329.

Taagepera, M., Noori, S., (2000), Mapping students' thinking patterns in learning organic chemistry by the use of Knowledge Space Theory. *J. Chem. Educ.*, **77**, 1224–1229.

Talanquer, V., (2013), When Atoms Want. *J. Chem. Educ.*, **90**, 1419–1424.

Tarhan, L., Sesen, B.A., (2012), Jigsaw cooperative learning: acid–base theories. *Chem. Educ. Res. Pract.*, **13**, 307–313.

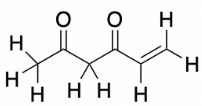
Wilson, J.W., (1994), Writing to Learn in an organic chemistry course. *J. Chem. Educ.*, **71**, 1019-1020.

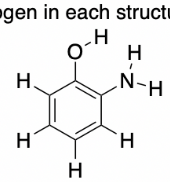
المنارة للاستشارات
www.manaraa.com

Appendix A
Supporting Information for Chapter 3

A.1 Examination items with acid–base themes

Problem A (4 points)
CIRCLE the most acidic hydrogen in each structure.

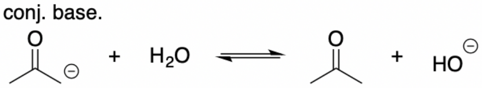




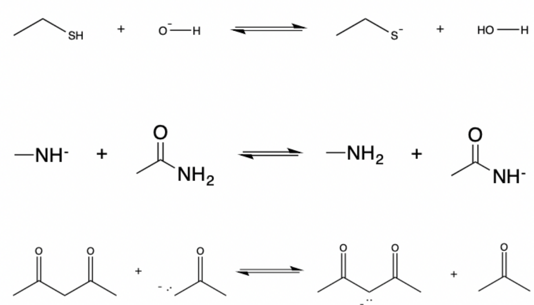
Problem B (10 points)
For each pair: (A). CIRCLE the most acidic structure. (B). Explain your reason.

Cmpd. 1	Cmpd. 2	Reason
CH_3CH_2SeH	CH_3CH_2SH	
CF_2HCH_2OH	CH_3CF_2OH	
$CH_3CH_2NH_2$	$CH_3CH_2CH_3$	
CH_2CH_2	C_2H_6	
CH_3COOH	CH_3CH_2CHO	

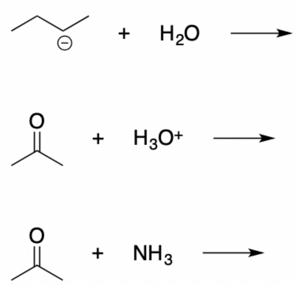
Problem C (6 points)
(A). Use arrows to show electron flow for this reaction. (B). Label the acid, base, conj. acid, and conj. base.



Problem D (6 points)
CIRCLE the side (products or reactants) of the equilibrium favored for these reactions.



Problem E (6 points)
Predict the product of the following Brønsted-Lowry or Lewis acid-base reactions.



www.manaraa.com

A.2 Categories included in logistic regression

Note About Categorization Scheme

Categories are not mutually exclusive. Common misspellings, abbreviations, and synonyms have been added as "synonyms" that are used in the terms/phrases as detected through the lexical analysis.

<Type>

Several words are added to a single type, so rules can be created. For example, the <attract> type includes the terms "attract", "attracted", "attracting", "attracts", "pulls", "pulling", "pulled", and "pull".

Rules

Rules involve a number of "any words" between given categories. This allows for patterns to be coded when there are additional words between the words of importance. The number of words varies based on what was observed in the training data set and is appropriate for each specific rule.

Type	Words included in type
<accept>	Accept, grab, get, steal, keep, pick up, receive, take, gain, acceptor, acceptance, attach, leave with, reach for
<acid>	HCl, H_3O^+, hydronium, hydrochloric acid
<act>	Act, acting, acts
<attract>	Pull, pulled, pulls, pulling, attract, attracted, attracting
<bond>	Bond, bonds
<break>	Break, breaks, breaking, broken
<charge>	Charge, charges, charged
<donate>	Push, share, let go, let go of, give up, donor, donation, lose, give, donate, transfer, taken from
<electrons>	Electrons, electron pair, bond electrons, bonding electrons
<form>	Form, formed, forming, make, making, made
<hydrogen>	Hydrogen, H
<ions>	Anions, ions, cations, ionized
<nucleophile/electrophile>	Nucleophile, electrophile
<oxygen>	Oxygen, O
<partial>	Slightly, slight, partially, partial, partial positive, partial negative
<strong>	Good, strong, strength
<water>	Water, H_2O

Category	*Use of Conjugate Acid/Base Terminology*
Description	Used the term "conjugate acid" and/or "conjugate base", or the word conjugate and either the word acid or base.
Terms/Phrases/Types	"Brønsted base", "Brønsted acid", "conjugate acid", "conjugate base"
Rules	contains words "conjugate" AND "base"; contains the words "conjugate" AND "acid"
Example Response(s)	"Then water produces its conjugate acid (hydronium ion) with a positive charge." "The bromine is attracting the electrons in the bond." "Water is acting as a Bronsted base by accepting the proton from the hydrochloric acid.""

المنارة للاستشارات
www.manaraa.com

Category	*Use of Electron Pairs Terminology*
Description	Used the term “electron pairs” and/or “pairs of electrons”.
Terms/Phrases/Types	“pair of electrons”, “electron pair”; *Note: “lone pair” is set as a synonym for “pair of electrons”*
Rules	(none)
Example Response(s)	“In this mechanism, the lone pairs on the oxygen bond to the proton connected to chlorine.” “The bond between the hydrochloric acid is breaking and the chlorine is taking the pair of electrons.”

Category	*Use of Electronegativity Terminology*
Description	Used the term “electronegativity”.
Terms/Phrases/Types	“Electronegativity”; *Note: The word “electronegative” has been assigned the synonym “electronegativity”.*
Rules	(none)
Example Response(s)	“The partially electronegative oxygen is being protonated by acquiring the H proton from the HCl.” “The electronegativity of oxygen is attracted to the positive charge of the hydrogen that is attached to the chlorine.”

Category	*Use of Forming Ions Process*
Description	Used the concept of formation of ions or the presence of ions in solution.
Terms/Phrases/Types	(none)
Rules	Contains the words “is” AND “ionized”; • Contains the words “ion” AND “solution”; <form> + <ions>: a word from the “form” type, 0 to 6, a word from the “ions” type
Example Response(s)	“Cl is able to form a stable ion due to its size and is easily removed.” “But hydrogen only wants to form a single bond so the hydrogen will break its bond with chlorine, making the chlorine an ion in the solution.”

Category	*Use of Nucleophile/Electrophile Terminology*
Description	Used nucleophile and/or electrophile terminology.
Terms/Phrases/Types	“Lewis acid”, “Lewis base”, “nucleophile”, “electrophile”
Rules	<act> + <nucleophile/electrophile>: a word from the type “act”, 0 to 6, a word from the type “nucleophile/electrophile;
Example Response(s)	“The water molecule is acting as a nucleophile and forms a bond with the hydrogen in hydrochloric acid.” “H+ is an electrophile accepting the free electrons from the polar water molecule.” “The H2O performs a nucleophilic attack on the HCl.”

Category	*Use of Accept Electrons Process*
Description	Used the concept of acceptance and the concept of electrons.
Terms/Phrases/Types	(none)
Rules	<accept> + <electrons>: a word from the type “accept”, 0 to 6, a word from the type “electrons”; <attract> + <electrons>: a word from the type “attract”, 0 to 6, a word from the type “electrons”; Contains the words “accept” AND “electrons”
Example Response(s)	“Cl (being a good leaving group) takes the electrons from the bond between H and Cl and leaves.” “The hydrochloric acid accepts the electrons and becomes an ion.”

المنارة للاستشارات
www.manaraa.com

Category	*Use of Accept Protons Process*
Description	Used the words acceptance and hydrogen but did not use the word electrons or electron pairs, or talked specifically about something accepting hydrogen.
Terms/Phrases/Types	“hydrogen acceptor”, “deprotonate”
Rules	<accept> + <hydrogen>: a word from the type “accept”, 0 to 6, a word from the type “hydrogen”; Contains the words “accept” AND “hydrogen” AND NOT “electrons” OR “electron pairs”
Example Response(s)	“Water is a weak base and will accept protons.” “So the water grabs the H and it becomes relatively stable.” H_2O looks to stabilize itself, so it takes the positive H.”

Category	*Use of Acid Strength Concept*
Description	Used the words “acid” and “strength”, or the phrases “hydrochloric acid” and “strong acid”.
Terms/Phrases/Types	(none)
Rules	Contains the words “strong” AND “acid”; Contains the phrases “hydrochloric acid” AND “strong acid”
Example Response(s)	“Hydrochloric acid is a very strong polar acid.” “HCl easily dissociates because it is a strong acid.”

Category	*Use of Attraction of Hydrogen Concept*
Description	Used the concept of attracting hydrogen.
Terms/Phrases/Types	(none)
Rules	<attract> + <hydrogen>: a word from the “attract” type, 0 to 6, a word from the “hydrogen” type.
Example Response(s)	“The partial negative side of oxygen in water will act as a base. This means it attacks the acid and attracts the partially positive H atom in HCl.”

Category	*Use of Bond Electrons Terminology*
Description	Used the phrase “bonding electrons” and/or used both the concept of bonds and the concept of electrons.
Terms/Phrases/Types	“bonding electrons”
Rules	Contains the words “bond” AND “electrons”
Example Response(s)	“This occurs because since chlorine is much more electronegative than hydrogen it is likely that it will break the bond and use those electrons to form a relatively stable ion.”

Category	*Use of Dissociation Process*
Description	Used the concept of dissociation.
Terms/Phrases/Types	(none)
Rules	Contains the words “hydrochloric acid” AND “dissociate”; Contains the words “dissociate” AND “completely”; Contains the words “bond” AND “dissociate”
Example Response(s)	“The chloride ion dissociates from hydrochloric acid.” “Strong acids dissociate completely when placed in water and forms ions.”

www.manaraa.com

Category	*Use of Donate Electrons Process*
Description	Used the words donation and electrons.
Terms/Phrases/Types	(none)
Rules	<electrons> + <donate>: a word from the type “electrons”, 0 to 6, a word from the type “donate”; Contains the words “donate” AND “electrons”
Example Response(s)	“The Cl breaks off from the HCl as the electrons are donated back to the Cl when the H attracts the oxygen.”

Category	*Use of Donate Protons Process*
Description	Used the concepts of donation and hydrogen but did not use the concepts of electrons or electron pairs, or talked specifically about something donating hydrogen.
Terms/Phrases/Types	“protonate”
Rules	<chemical> + <donate> + <hydrogen>: a word from the type “chemical”, 0 to 6, a word from the type “donate”, 0 to 6, a word from the type “hydrogen”; Contains the words “donate” AND “hydrogen”, AND NOT “electrons” OR “electron pair”
Example Response(s)	“HCl is donating its hydrogen to the H_2O.” “Hydrochloric acid is acting as the Bronsted acid by donating a proton to the Bronsted base.”

Category	*Use of Sharing Electrons Process*
Description	Used the concept of sharing electrons.
Terms/Phrases/Types	“sharing electrons”
Rules	Contains the words “sharing” AND “electrons”
Example Response(s)	“The Cl will take the electrons it was sharing with the H.”

Category	*Use of Hydrogen Actions Process*
Description	Used the concept of hydrogen performing a task, such as being donated or bonding to something.
Terms/Phrases/Types	“hydrogen bonds” *Note: students very rarely used this to refer to actual hydrogen bonding. Instead, this phrase was used to say that the atom hydrogen was forming a bond with something.*
Rules	<hydrogen> + <donate>: a word from the type “hydrogen”, 0 to 6, a word from the type “donate”
Example Response(s)	“The hydrogen bonds with the oxygen to create H_3O^+.” “This reaction takes the hydrogen from the HCl and attaches it to the water as the bond between HCl is broken.”

Category	*Use of Hydrogen and not Electrons Terminology*
Description	Used the word hydrogen or proton but has no mention of electrons or electron pairs.
Terms/Phrases/Types	(none)
Rules	Contains the word “hydrogen”, AND NOT “electrons” OR “electron pair”
Example Response(s)	“The hydrogen is breaking a bond with chlorine and making a bond with oxygen.”

المنارة للاستشارات
www.manaraa.com

Category	*Use of Oxygen as an Electron Donor Process*
Description	Used the concept of water or oxygen from water as an electron donor.
Terms/Phrases/Types	“water donate”
Rules	<water> + <donate> + <electrons>: a word from the type “water”, 0 to 6, a word from the type “donate”, 0 to 6, a word from the type “electrons”; Contains the words “oxygen” AND “donate” AND “electrons”
Example Response(s)	“The water shares its electrons with the hydrogen, forming a new bond.” “The Cl breaks off from the HCl as the electrons are donated back to the Cl when the H attracts the oxygen.”

Category	*Use of Partial Charges Concept*
Description	Used the word “partial” and either “positive” or “negative”.
Terms/Phrases/Types	“partial negative charge”, “partial positive charge”
Rules	<partial> + <charge>: a word from the type “partial”, 0 to 6, a word from the type “charge”; Contains the words “partial” AND “positive”; Contains the words “partial” AND “negative”
Example Response(s)	“The partial negative charge of water is attracting the partial positive charge of the hydrogen.”

A.3 Categories not included in logistic regression

Category	*Use of Bond Breaking Process*
Description	Used the concept of a bond breaking.
Terms/Phrases/Types	(none)
Rules	<break> + <bond>: a word from the type “break”, 0 to 6, a word from the type “bond”
Example Response(s)	“The O is taking the H which breaks the bond with the Cl and the Cl leaves as the leaving group.”

Category	*Use of Bond Forming Process*
Description	Used to concept of a bond forming.
Terms/Phrases/Types	“new bond”
Rules	<bond> + <form>: a word from the type “bond”, 0 to 6, a word from the type “form”
Example Response(s)	“The oxygen has 2 lone pairs so it wants to form a new bond.” “What you have is the H atom forming a bond with water to form the hydronium ion.”

Category	*Use of Charges Terminology*
Description	Used terminology that relates to charges.
Terms/Phrases/Types	“carbocation”, “formal charge”, “positive charge”, “negative charge”
Rules	(none)
Example Response(s)	“The Cl dissociates from the H because it is stable with a negative charge.” “The chloride is gaining a lone pair making the formal charge negative.”

www.manaraa.com

Category	*Use of Hydrochloric Acid Terminology*
Description	Used the term “hydrochloric acid”.
Terms/Phrases/Types	“hydrochloric acid”
Rules	(none)
Example Response(s)	“The hydrochloric acid will readily dissociate as a strong acid.”

Category	*Use of Leaving Group Terminology*
Description	Used terms that relate to leaving groups or chlorine leaving.
Terms/Phrases/Types	“chlorine leave”, “leaving group”, “good leaving group”
Rules	(none)
Example Response(s)	“Being a good leaving group, Cl is able to leave and be stable on its own.” “Hydrochloric acid dissociates when in water, making the chlorine leave.”

Category	*Use of Octet Terminology*
Description	Used the word “octet” or the phrase “octet rule”.
Terms/Phrases/Types	“octet”, “octet rule”
Rules	(none)
Example Response(s)	“Halogens readily accept hydrogens in order to fill their outermost valence shell and fulfill the octet rule.”

Category	*Use of Polarity Concept*
Description	Used the concept of a polar bond or a polar molecule.
Terms/Phrases/Types	“polar molecule”
Rules	Contains the words “polar” AND “bond”
Example Response(s)	“This occurs because water is a polar molecule.” “Oxygen is very negative for which the O-H bond is very polar.”

Category	*Use of Stability Concept*
Description	Used terms that relate to stability.
Terms/Phrases/Types	“stability”, “stable anion”
Rules	(none)
Example Response(s)	“The reaction occurs because Cl is a very stable anion.” “This reaction occurs in order for there to be stability between the acids and the bases.”

Category	*Use of Water Actions Process*
Description	Used words that imply water completing an action.
Terms/Phrases/Types	“water gains”, “water molecule gains”, “water attaches”, “water attacks”
Rules	(none)
Example Response(s)	“The Oxygen in the water attacks the hydrogen and the bond between the Hydrogen and Chloride breaks.” “The nearby water attaches to the free hydrogen thus creating a hydronium.”

Category	*Use of Oxygen Bonds Process*
Description	Used the concept of oxygen bonding.
Terms/Phrases/Types	(none)
Rules	<oxygen> + <bond>: a word from the type “water”, 0 to 6, a word from the type “bond”
Example Response(s)	“The electrons from water are making a bond with hydrogen.”

المنارة للاستشارات
www.manaraa.com

Appendix B

Supporting Information for Chapter 5

B.1 Categories used in logistic regression

Categories are not mutually exclusive.

<type>: Several words (or just one word) can be added to a type so that rules can be created. A list of types can be found below.

Rules allow for six "any words" between types called for in the rule.

Category	*Absence of explanation*
Description	Response does not include terminology and phrases associated with the *why* of the reaction.
Terms/Phrases/Types	<absence of explanation>
Rules	Does NOT include: <attract> <electronegativity> <partial> <stability> <degree> <strong> <weak> "excess"
Example Response(s)	"A carbocation forms where bromine left the substrate. The bromine leaves the substrate and the nucleophile attacks."

Category	*Accept/donate electrons*
Description	Response talks about accepting or donating electrons
Terms/Phrases/Types	None
Rules	<donate>+ <electrons> <accept>+ <electrons> <lose>+ <electrons> <keep>+ <electrons>
Example Response(s)	"…creates a positive carbocation which in turn accepts a pair of electrons from the alcohol." "The oxygen molecule comes in and donates its electrons." "The oxygen keeps the electrons that were forming the bond." "Carbon and bromine are breaking to give those two electrons to bromine."

المنارة للاستشارات
www.manaraa.com

Category	*Attraction*
Description	Response talks about the attraction of something.
Terms/Phrases/Types	None
Rules	<attract> + <negative> <carbocation> + <attract> <opposites> + <attract> <attract> + <nucleophile> <electrons> + <attract> <positive> + <attract> <negative> + <attract> <attract> + <positive> <attract> + <electrons> <starting material> + <attract> <attract> + <starting material> <attract> + <bond> <bond> + <attract> <attract> + <carbocation> <nucleophile> + <attract>
Example Response(s)	"…to form a tertiary carbocation, which then attracts the oxygen molecule from the alcohol." "With ethanol acting as a nucleophile it is attracted to the carbocation and bonds to the structure." "The bromine is attracting the electrons in the bond."

Category	*Bond breaks*
Description	Response talks about the breaking of a bond.
Terms/Phrases/Types	None
Rules	<bond> + <break> <break> + <bond>
Example Response(s)	"Bromide's bond to carbon is being broken." "The ethanol comes back around and attacks the hydrogen breaking the bond and forming" t-butyl ether."

Category	*Bond electrons*
Description	Response talks about the electrons in the bond or electrons bonding.
Terms/Phrases/Types	None
Rules	<bond> + <electrons> <electrons> + <bond>
Example Response(s)	"The extra pair of electrons will bond to the positive charge." "The bromine is attracting the electrons in the bond." "The bond electrons stay with the oxygen"

Category	*Bond forms*
Description	Response talks about a bond forming.
Terms/Phrases/Types	"new bond"
Rules	<bond> + <form> <form> + <bond>
Example Response(s)	"A bond forms between the carbocation and the oxygen of the ethanol." "One of the lone pairs from the alcohol group attached to the ethanol molecule attacks the carbocation and forms a bond with it." "The ethanol comes in and attacks the positive carbocation forming a new bond."

المنارة للاستشارات
www.manaraa.com

Category	*Carbocation*
Description	Response includes the carbocation type.
Terms/Phrases/Types	<carbocation>
Rules	None
Example Response(s)	“A bromine takes the electrons in its bond, leaving a carbocation.”

Category	*Carbocation attack*
Description	Response talks about the carbocation being attacked or the carbocation attacking.
Terms/Phrases/Types	None
Rules	<attack> + <carbocation> <attack> + <positive> <bond> + <carbocation> <carbocation> + <accept> <attack> + <charge> <carbocation> + <attack>
Example Response(s)	“The alcohol attacks because the C with the cation needs to be stabilized.” “. . . leaving a tertiary carbocation which is then attacked by the ethyl alcohol.” “Ethanol is used to make the new structure because it will bond to the carbocation.” “. . . creates a positive carbocation which in turn accepts a pair of electrons from the alcohol.”

Category	*Carbocation degree*
Description	Response talks about the degree of the carbocation (i.e., tertiary).
Terms/Phrases/Types	None
Rules	<degree> + <carbocation> <carbocation> + <degree>
Example Response(s)	“Bromine leaves, leaving a tertiary carbocation.” “The reaction is S_N1 because the carbocation is tertiary.”

Category	*Deprotonate*
Description	Response talks about deprotonation or accepting a proton.
Terms/Phrases/Types	<deprotonate>
Rules	<accept> + <hydrogen> <hydrogen> + <accept>
Example Response(s)	“A second alcohol comes in to deprotonate the substrate.” “The hydrogen is attacked by the oxygen to remove the hydrogen.”

Category	*Donate hydrogen*
Description	Response talks about donating a proton.
Terms/Phrases/Types	“protonate”
Rules	<donate> + <hydrogen> <lose> + <hydrogen> <hydrogen> + <donate>
Example Response(s)	“A proton is donated from the protonated alcohol.” “Water is playing the role of a base, therefore protonating the alcohol that was there.”

www.manaraa.com

Category	*“Don’t know”*
Description	Response states a phrase related to “I don’t know”
Terms/Phrases/Types	<do not know>
Rules	None
Example Response(s)	“I am unable to answer the question to the desired depth.” “I’m not sure if this is in depth enough.”

Category	*Electron attack*
Description	Response talks about electrons attacking or being used to form a bond.
Terms/Phrases/Types	None
Rules	<electrons> + <attack> <attack> + <electrons>
Example Response(s)	“OH is a good nucleophile because it has a pair of electrons it can attack with.” “An EtOH lone pair then attacks the carbocation’s positive charge.” “It takes the electrons in the bond that was once formed with the substrate.”

Category	*Electron terminology*
Description	Response uses electron terminology.
Terms/Phrases/Types	<electrons>
Rules	None
Example Response(s)	“The C-Br bond will break giving its electrons to Br.” “Then the oxygen from EtOH has lone pairs that will attack the carbocation.”

Category	*Electronegativity*
Description	Response invokes the concept of electronegativity.
Terms/Phrases/Types	<electronegativity>
Rules	<sharet> + <electrons> <electrons> + <density> <density> + <electrons> <needs> + <electrons> <wants> + <electrons> <electrons> + <sharing> <attract> + <electrons> <electrons> + <attract>
Example Response(s)	“The lone pairs in the oxygen are attracted to the carbocation.” “. . . it needs more electrons due to it having a cation.” “Since it is more electronegative than its substrate it takes the electrons in the bond.” “Br is taking the shared electrons.” “Different areas of electron density cause a shift in reactants to form new products.”

Category	*Electrophile accepts*
Description	Response talks about the electrophile accepting electrons.
Terms/Phrases/Types	None
Rules	<electrophile> + <accept> <carbocation> + <accept>
Example Response(s)	“An electrophile accepts a pair of electrons to form a new covalent bond.” “The tertiary carbocation accepts the electrons from the OH.”

المنارة للاستشارات
www.manaraa.com

Category	*Eliminate charge*
Description	Response talks about neutralizing the charge.
Terms/Phrases/Types	<neutralize>
Rules	<stabilize> + <charge> <lose> + <charge> <eliminate> + <charge>
Example Response(s)	“The remaining reaction occurs to remove the charge on water and balance the molecule.” “Ethanol is in excess so it will need to be used again to neutralize the alcohol group.” “The bond with hydrogen is broken so the oxygen can return to a neutral charge.” “Another ethanol molecule comes along and remove the hydrogen from the positively charged oxygen in order to eliminate the positive charge.” “Bromine is a very good leaving group (it is able to stabilize a negative charge).”

Category	*Good leaving group*
Description	Response describes a good leaving group.
Terms/Phrases/Types	None
Rules	<good> + <leaving group> <wants to> + <leave>
Example Response(s)	“This reaction occurs because bromine is a very good leaving group.” “The reaction occurs because the tertiary bromine wants to leave as a leaving group.”

Category	*Leaving group leaves*
Description	Response talks about the leaving group leaving.
Terms/Phrases/Types	None
Rules	<leaving group> + <leave> <halogen> + <leave>
Example Response(s)	“This reaction occurs because the leaving group leaves.” “The nucleophile has to wait for the bromine to leave in order to bond to the positive carbocation.”

Category	*Nucleophile/electrophile terminology*
Description	Response talks about a nucleophile or an electrophile.
Terms/Phrases/Types	<nucleophile>, <electrophile>
Rules	None
Example Response(s)	“The nucleophile attacks the electrophile.”

المنارة للاستشارات
www.manaraa.com

Category	*Opposites*
Description	Response talks about opposing charges.
Terms/Phrases/Types	“dipole”
Rules	<opposite> + <charge> <positive> + <negative> <negative> + <positive> <negative> + <attack> + <positive> <positive> + <attack> + <negative>
Example Response(s)	“Opposite charges attract and like charges repel.” “The carbocation is positively charged attracting the partially negative oxygen in ethanol.” “OH then attacks the carbocation because a negative charge is attracted to the positive charge.” “The second step would be an O which maintains a partial negative charge attacking the positive compound.” “A positive dipole is created on the oxygen of ethanol so another ethanol can remove the hydrogen.”

Category	*Partial charges*
Description	Response talks about partial charges.
Terms/Phrases/Types	None
Rules	<partial> + <charge>
Example Response(s)	“Ethanol acts as a nucleophile because of the partial negative charge on oxygen.”

Category	*Reaction terminology*
Description	Response labels the type of reaction (e.g., S_N1).
Terms/Phrases/Types	<reaction type>
Rules	None
Example Response(s)	“Since it is an S_N1 reaction, the leaving group leaves first.” “S_N1 is a unimolecular substitution reaction.” “There is also not heat present so the reaction will not favor elimination.”

Category	*Solvent terminology*
Description	Response talks about the solvent (e.g., ethanol) or solvent type (e.g., protic).
Terms/Phrases/Types	<solvent>
Rules	None
Example Response(s)	“The reaction occurs because the substrate is tertiary and the solvent is polar protic.” “In the presence of polar aprotic solvents, the leaving group leaves and forms a carbocation.”

Category	*Stability of carbocation*
Description	Response talks about the stability of the carbocation.
Terms/Phrases/Types	None
Rules	<carbocation> + <stability> <stability> + <carbocation> <carbcation> + <deficient>
Example Response(s)	“The carbocation is unstable and needs to acquire electrons to stabilize the charge.” “After bromine leaves, a stable tertiary carbocation forms in its place.” “A tertiary carbocation forms that is deficient in electrons.”

المنارة للاستشارات
www.manaraa.com

Category	*Stability terminology*
Description	Response includes terminology related to stability.
Terms/Phrases/Types	<stability>
Rules	None
Example Response(s)	"This reaction occurs in order to achieve a more stable molecule." "Another ethanol molecule can further stabilize the reaction by removing the hydrogen from the first oxygen." "This reaction occurs due to stability." "The positive charge is not stable."

Category	*Sterics*
Description	Response includes terminology related to sterics.
Terms/Phrases/Types	<sterics>
Rules	None
Example Response(s)	"Since the substrate is a tertiary carbocation, it is considered sterically hindered." "This occurs because the reaction consists of a bulky base." "The alkyl halide leaves first to create a carbocation and to reduce crowdedness."

Category	*Temperature*
Description	Response includes terminology related to temperature.
Terms/Phrases/Types	<temperature>
Rules	None
Example Response(s)	"Since it is tertiary and no heat or strong base, it prefers S_N1." "The reaction occurs because of lower temperature.

Category	*"Wants"*
Description	Response talks about molecules "wanting" to do things.
Terms/Phrases/Types	<wants to>
Rules	None
Example Response(s)	"This reaction occurs because bromine is a leaving group that generally wants to break a bond."

Category	*Weak/strong base*
Description	Response describes the strength of the base.
Terms/Phrases/Types	None
Rules	<weak> + <base> <strong> + <base> <base> + <weak> <base> + <strong>
Example Response(s)	"S_N1 favors weak bases." "This happens because OH is a strong base." "The leaving group leaves because the substrate is tertiary and the base is weak."

المنارة للاستشارات
www.manaraa.com

Category	*Weak/strong nucleophile*
Description	Response describes the strength of the nucleophile.
Terms/Phrases/Types	None
Rules	<weak> + <nucleophile> <strong> + <nucleophile> <nucleophile> + <weak> <nucleophile> + <strong>
Example Response(s)	"S_N1 has a weak nucleophile with a highly substituted leaving group." "The reaction is S_N1 because it is a tertiary substrate/electrophile and the nucleophile is a strong nucleophile." "Since the nucleophile is weak it then becomes an S_N1 mechanism." "The leaving group and nucleophile are strong."

B.2 Types used in logistic regression

Type	Words included in type
Accept	Accept, attach, gain, get, keep, leave with, pick up, pull off, reach, receive, steal, swipe, interact, draw
Acid	acid
Act	act
Attack	Attack, grab, take
Attract	Attract, draw, interact, gravitate, intermolecular force, pull, go to
Base	Base
Because	Because, due to, since
Bond	bond
Break	Break, disconnect
Carbocation	Positive carbon, cation, carbocation
Carbon	Carbon
Charge	Charge
Degree	Degree, tertiary, quaternary
Density	Rich, poor, density, deficient, sufficient, lack
Deprotonate	Deprotonate, remove proton, take proton
Do not know	Don't know, don't remember, no idea, not exactly sure, not sure, unable to answer, unknown, not too sure
Donate	Donate, give, let go, take from, transfer
Electronegativity	Electronegative, electropositive
Electrons	Bond electrons, electron cloud, electron pair, electrons, lone pair
Electrophile	electrophile
Form	Form, make, produce
Halogen	Bromine, chlorine, iodine, fluorine, halogen
Hydrogen	Hydrogen, proton
Ions	Ion, anion, cation
Join	Join, come together
Just because	Appropriate, conditions, how it works, nature of reaction, just because
Leave	Alone, depart, detach, exit, go away, leave, move, off by itself, separate, break, kick off, remove itself

المنارة للاستشارات
www.manaraa.com

Type	Words included in type
Leaving group	Leaving group, lg
Lose	lose
Need	Need, require
Negative	Negative, minus
Neutralize	Cancel out, get rid of charge, get rid of the charge, neutralize, no charge, no net charge, not charged, remove charge, remove the charge, uncharge, eliminate charge, eliminate the charge, neutral, relieve
Nucleophile	nucleophile
Number of steps	Step, more than one, multiple
Opposite	Opposite, different, unbalance, unequal
Partial	Partial, slight, unequal
Positive	Cation, lack, not neutral, plus, positive
Preposition	Between, from, with
Product	Product, final
Reaction type	Addition, rearrangement, eliminate, substitute, condition, e1, e2, sn1, sn2, backside attack
Sharing	Share, divide, distribute
Size	Small, large, big, tiny, size
Solvent	Excess, polar, protic, aprotic, solvent, nonpolar
Stability	Balance, stable, unstable
Starting material	Original compound, starting material, reactant, reagent, substrate, ethanol, methanol, alcohol, tert-butyl bromine, tert-butyl iodine, tert-butyl chlorine, water, oxygen
Sterics	Steric, bulky, hinder, crowd, constraint, inaccessible
Strong	Fine, good, strong, excellent
Temperature	Cold, cool, heat, hot, warm, temperature
Wants to	Want, fine with, happy to, like, prefer, ready to, want, desire
Weak	Weak, not strong

المنارة للاستشارات
www.manaraa.com

Appendix C

Publisher permissions documentation

C.1 Chapter 3 (Journal of Chemical Education)

Copyright Clearance Center RightsLink®

Home | Create Account | Help | LIVE CHAT

ACS Publications
Most Trusted. Most Cited. Most Read.

Title: Using Lexical Analysis To Predict Lewis Acid–Base Model Use in Responses to an Acid–Base Proton-Transfer Reaction

Author: Amber J. Dood, Kimberly B. Fields, Jeffrey R. Raker

Publication: Journal of Chemical Education

Publisher: American Chemical Society

Date: Aug 1, 2018

Copyright © 2018, American Chemical Society

LOGIN

If you're a copyright.com user, you can login to RightsLink using your copyright.com credentials.

Already **a RightsLink user** or want to learn more?

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

BACK | CLOSE WINDOW

Copyright © 2019 Copyright Clearance Center, Inc. All Rights Reserved. Privacy statement. Terms and Conditions.
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

المنارة للاستشارات
www.manaraa.com

C.2 Chapter 4 (Canadian Journal of Chemistry)

Canadian Science Publishing LICENSE
TERMS AND CONDITIONS

Sep 25, 2019

This is a License Agreement between Univeristy of South Florida -- Amber Dood ("You") and Canadian Science Publishing ("Canadian Science Publishing") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Canadian Science Publishing, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	4676100563091
License date	Sep 25, 2019
Licensed content publisher	Canadian Science Publishing
Licensed content title	Canadian journal of chemistry
Licensed content date	Jan 1, 1951
Type of Use	Thesis/Dissertation
Requestor type	Author of requested content
Format	Print, Electronic
Portion	chapter/article
The requesting person/organization is:	Amber J. Dood
Title or numeric reference of the portion(s)	Entirety of article
Title of the article or chapter the portion is from	Development and evaluation of a Lewis acid–base tutorial for use in postsecondary organic chemistry courses
Editor of portion(s)	N/A
Author of portion(s)	Dood, Amber J.; et al
Volume of serial or monograph.	N/A
Page range of the portion	
Publication date of portion	Jul 2, 2019
Rights for	Main product and any product related to main product
Duration of use	Life of current edition
Creation of copies for the disabled	no
With minor editing privileges	no
For distribution to	United States
In the following language(s)	Original language of publication
With incidental promotional use	no
The lifetime unit quantity of new product	Up to 499
Title	Development and evaluation of a Lewis acid–base tutorial for use in postsecondary organic chemistry courses
Institution name	University of South Florida
Expected presentation date	Mar 2020
Total (may include CCC user fee)	0.00 USD
Terms and Conditions	

المنارة للاستشارات
www.manaraa.com

TERMS AND CONDITIONS

The following terms are individual to this publisher:

None

Other Terms and Conditions:

STANDARD TERMS AND CONDITIONS

1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the "Work(s)"). Copyright Clearance Center, Inc. ("CCC") grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the "Rightsholder"). "Republication", as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. "User", as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a "freelancer" or other third party independent of User and CCC, such party shall be deemed jointly a "User" for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.

3. Scope of License; Limitations and Obligations.

3.1 All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.

3.2 General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, 29118 Network Place, Chicago, IL 60673-1291. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on "net 30" terms. While User may exercise the rights licensed immediately upon issuance of the Order Confirmation, the license is automatically

المنارة للاستشارات
www.manaraa.com

revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.

3.3 Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is "one-time" (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User's stock at the end of such period).

3.4 In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.

3.5 Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Republished with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc. " Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

3.6 User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

4. Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.

5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK,

المنارة للاستشارات
www.manaraa.com

EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.

6. Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS". CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

7. Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

8. **Miscellaneous.**

8.1 User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.

8.2 Use of User-related information collected through the Service is governed by CCC's privacy policy, available online here:
http://www.copyright.com/content/cc3/en/tools/footer/privacypolicy.html.

8.3 The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in the new material which includes the Work(s) licensed under this Service.

8.4 No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order

المنارة للاستشارات
www.manaraa.com

Confirmation and/or in these terms and conditions or CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

8.5 The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-750-8400 or send an e-mail to info@copyright.com.

v 1.1

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

المنارة للاستشارات

www.manaraa.com

C.3 Chapter 5 (Chemistry Education Research and Practice)

Analyzing explanations of substitution reactions using lexical analysis and logistic regression techniques

A. J. Dood, J. C. Dood, D. Cruz-Ramírez de Arellano, K. B. Fields and J. R. Raker, *Chem. Educ. Res. Pract.*, 2020, **21**, 267
DOI: 10.1039/C9RP00148D

If you are not the author of this article and you wish to reproduce material from it in a third party non-RSC publication you must formally request permission using Copyright Clearance Center. Go to our Instructions for using Copyright Clearance Center page for details.

Authors contributing to RSC publications (journal articles, books or book chapters) do not need to formally request permission to reproduce material contained in this article provided that the correct acknowledgement is given with the reproduced material.

Reproduced material should be attributed as follows:

- For reproduction of material from NJC:
  Reproduced from Ref. XX with permission from the Centre National de la Recherche Scientifique (CNRS) and The Royal Society of Chemistry.
- For reproduction of material from PCCP:
  Reproduced from Ref. XX with permission from the PCCP Owner Societies.
- For reproduction of material from PPS:
  Reproduced from Ref. XX with permission from the European Society for Photobiology, the European Photochemistry Association, and The Royal Society of Chemistry.
- For reproduction of material from all other RSC journals and books:
  Reproduced from Ref. XX with permission from The Royal Society of Chemistry.

If the material has been adapted instead of reproduced from the original RSC publication "Reproduced from" can be substituted with "Adapted from".

In all cases the Ref. XX is the XXth reference in the list of references.

If you are the author of this article you do not need to formally request permission to reproduce figures, diagrams etc. contained in this article in third party publications or in a thesis or dissertation provided that the correct acknowledgement is given with the reproduced material.

Reproduced material should be attributed as follows:

- For reproduction of material from NJC:
  [Original citation] - Reproduced by permission of The Royal Society of Chemistry (RSC) on behalf of the Centre National de la Recherche Scientifique (CNRS) and the RSC
- For reproduction of material from PCCP:
  [Original citation] - Reproduced by permission of the PCCP Owner Societies
- For reproduction of material from PPS:
  [Original citation] - Reproduced by permission of The Royal Society of Chemistry (RSC) on behalf of the European Society for Photobiology, the European Photochemistry Association, and RSC
- For reproduction of material from all other RSC journals:
  [Original citation] - Reproduced by permission of The Royal Society of Chemistry

If you are the author of this article you still need to obtain permission to reproduce the whole article in a third party publication with the exception of reproduction of the whole article in a thesis or dissertation.

Information about reproducing material from RSC articles with different licences is available on our Permission Requests page.

المنارة للاستشارات
www.manaraa.com