Spring 2024

Characterizing and Developing Chemistry Students' Data Analysis and Interpretation of Chemical Data

Stephanie A. Berg
University of Nebraska-Lincoln

CHARACTERIZING AND DEVELOPING CHEMISTRY STUDENTS' DATA ANALYSIS AND INTERPRETATION OF CHEMICAL DATA

by

Stephanie A. Berg

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Chemistry

Under the Supervision of Professor Alena Moon

Lincoln, Nebraska

May, 2024

CHARACTERIZING AND DEVELOPING CHEMISTRY STUDENTS' DATA ANALYSIS AND INTERPRETATION OF CHEMICAL DATA

Stephanie A. Berg, Ph.D.

University of Nebraska, 2024

Advisor: Alena Moon

All chemistry students must develop competency in analyzing and making sense of data. However, there are many difficulties that chemistry students may experience while analyzing data. Many students may not use relevant prior knowledge to aid in making sense of data, or they may not form conclusions using all the data provided. Additionally, prior knowledge seems to influence one's data analysis, but little is known about how students use it to make sense of data. Thus, I interviewed undergraduate students as they analyzed graphical data for a task and characterized how they used their prior knowledge throughout. My findings suggest that students' prior knowledge helped to form a frame for students. This frame is then used throughout students' sensemaking to help search for and identify relevant data and evaluate data against their frame to aid in decision-making. Because there are limited classroom interventions designed to help develop undergraduate students' data analysis competencies, I designed a study in which undergraduate students compared their data analyses to pre-constructed sample responses via a simulated peer review. My findings suggest that providing students with the opportunity to compare their analyses against other responses, practice giving feedback, and reflect on their work may provide opportunities to generate internal feedback on their performance. Depending on the nature of the internal feedback (i.e., if it is critical or not), students may revise to improve their work. Finally, as part of contributing new knowledge to their field, chemistry graduate students must learn how to best respond to data that is discrepant with their expectations. Yet, there is little research on how chemistry graduate students analyze data, and none that explores how they respond to unexpected data. For this reason, I interviewed chemistry graduate students as they analyzed multiple data sets to explain a chemical phenomenon, and I characterized how students responded to unexpected data using Data-Frame Theory. My findings indicate that students respond to discrepant data in several ways, and each response is capable of progressing students' sensemaking to achieve the goals of the analysis.

REQUIRED STATEMENT

I certify that all of the work described within this dissertation is the original work of the author. Any published (or unpublished) ideas and/or techniques from the work of others are fully acknowledged in accordance with the standard referencing practices.

ACKNOWLEDGEMENTS

They say that it takes a village to raise a child, and I would argue that it also takes a village to raise a PhD. I would not have been able to do any of this without the help of my family, friends, and colleagues.

First, I would not have been able to pursue and sustain my education without the support of my family. Mom and Dad, I can't thank you enough for the countless times you have taken Casper and me to different museums and our many family visits to Half-Price Books. Our weekly phone calls and texts have kept me grounded and refreshed going into the coming weeks. Casper, your messages are always a bright spot in my day, and I cherish every picture of the rats you send. Let's play Nancy Drew soon. You all make me proud to be the shortest member of the Berg family. I love you dearly.

An enormous thank you to all my friends inside and outside of CER. Celeste Goodwin, our friendship means the world to me, and without our Discord chats, I would not have finished any of my dishes. Bud "Bestie" Jenkins, our DnD sessions and fun outings have truly brought so much queer joy to my life in Nebraska. Austin Aguilera, I love discussing different music and movies with you! Crystal Uminski, your dedication to good research and amazingly coordinated outfits and presentations make you a force to reckon with. Samantha Houchlei, who'd have thought two queer brunettes with glasses in CER would get along so well? Keenan Noyes did! Samantha and Keenan, you two have been incredible peers and mentors in the CER community. Alisha Szozda, Bri Martinez, and Melissa Collini, I cannot thank you all enough for meeting weekly for our goals group where we had such great talks. Archer Harrold, Rebecca Fantone, and Ina Zaimi, you all are such amazing researchers and I look forward to countless more talks and catchups! Katherine Johnson, I genuinely would not have survived Quantum without you; you are going to be an incredible mom with the smartest child out there. Erika De Leon, grading organic exams would have been much more painful without you hyperfocusing with me, and I am ever grateful for our hallway chats. Mary Tess Urbanek, not only do I get to call you a labmate, but I also get to call you a good friend!

Speaking of labmates… Haiyan Al Fulaiti, Matthew Breuer, Mary Tess Urbanek, John Zhou, and Leah Zohner, you have all helped provide such a solid space to be a researcher. You have all taught me so much, whether it's research-related like how to be an effective mentor, or non-research things like how important it is to wash your rice. I am quite proud of all of you, and I cannot see what will come for you in the next few years. Thank you all for making such a fantastic Moon Group to work with every day.

There would be no Moon Group without my advisor, Dr. Alena Moon. Alena, you have taught me so much in the last five years. You have truly helped me develop into the researcher I am today. You have given me the freedom and space to grow and become an independent researcher. I would not have the appreciation for a quality research design, a clearly written methods section, and a strong argument without your guidance. You have also helped guide me to be a better person and ally, and I cannot thank you enough for that. I am ecstatic to see what is in store for you in North Texas.

To my committee members, thank you all so much for your guidance and encouragement throughout the years. Dr. Sinitskii, the kindness you showed in my first year of graduate school did not go unnoticed, and I likely would not have continued without it. Dr. Cheung, your encouragement in pursuing the graduate writing retreat has helped strengthen my writing skills immensely. Dr. Couch, your thoughtful questions and support in pursuing the next steps in my career have helped so much. Dr. Li, I have such an immense appreciation for computational chemistry and quantum mechanics after taking courses with you. Thank you all for encouraging me and pushing me to become a better chemistry education researcher.

To my professors and teachers throughout K-12 to graduate school, thank you so much for helping me to continue growing as a learner. Mrs. Wahlin-Fiskum, Sr. Boeckers, Sra. Eiynck, and Mr. Manninen, thank you all so much for helping me grow in my critical reflection and language skills. To my professors at the University of Minnesota, Morris, I would not be half the student I am today without you. Thank you especially to Dr. Abbey Fischer, Dr. Nancy Carpenter, Dr. Ted Pappenfus, Dr. Zack Mensinger, Dr. Bryan Nell, Dr. Kara Nell, Dr. Jenn Goodnough, and Dr. Joe Alia for all creating such a welcoming and student-centered chemistry department. You all helped make me the chemist I am today.

I cannot express enough gratitude to those who have supported me in my mental health throughout my graduate school experience. I would like to especially thank my PA, Kayla Fink, for always working with me to find the medications to help me be a functioning person. I also want to thank my counselor, Jake, for guiding me in learning self-compassion, boundaries, and advocating for myself.

My wonderful partner, Emma, I cannot thank you enough for all the support that you have given me in our time together, whether it's going for walks or helping me come up with dinner ideas. I only hope I can show you an ounce of the same encouragement when it comes time for your thesis defense. I love you, HERB, and I am so sorry I am defending around the same time as our anniversary. I will make up for it next year!

Finally, I cannot express my gratitude to my cat, Franklin Berg enough. His near-daily cuddles and comical demands for lettuce, fish, and Tractor Supply Company wet food have truly

kept me grounded during this time. I promise I will eventually buy you that new cat tree so that you can watch all the creatures and greenery outside.

TABLE OF CONTENTS

Page

LIST of MULTIMEDIA OBJECTS

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Students of today are exposed to different sources of information and evidence that influence their decision-making. Many of these sources of information may be false (e.g., mis- and dis-information) or manipulated in some way (e.g., mal-information), and ultimately can cause real harm when used as the basis for decisions (Carmi et al., 2020). To prepare students to engage with such information in their everyday lives, some experts have argued for further developing students' understanding of science practices (Sharon & Baram-Tsabari, 2020). This echoes chemistry educators' calls for reforms to increase student engagement in science practices in post-secondary STEM courses (Cooper et al., 2015; National Research Council, 2012; Talanquer & Pollard, 2010).

Science practices consist of the methods that scientists use to "do" their science. One essential science practice for STEM students to develop competency in is data analysis. As undergraduate and graduate students progress through their education, they must continue building competency in data analysis to be prepared for the STEM workforce (Cooper et al., 2015; National Academies of Sciences, 2018). Although the Next Generation Science Standards has outlined the data analysis competencies students should be able to demonstrate at a variety of grade levels (National Research Council, 2012), there is a limited understanding of how this competency develops past the K-12 level. Thus, there is a need to further investigate how chemistry students engage in making sense of data at the postsecondary level and beyond.

Data Analysis and Graph Interpretation

One of the most common forms of data that students will encounter and analyze is the graph. The steps associated with data analysis and graph interpretation are comparable and share similar challenges. To begin, an individual will first encode the features of the data representation and identify the relevant features that are important for the remainder of the analysis (Carpenter & Shah, 1998; Glazer, 2011; Shah & Hoeffner, 2002; Zagallo et al., 2016). The encoded features can then be used to identify any relevant patterns or relationships that may be present (Ratwani et al., 2008; Zagallo et al., 2016). During these steps, many encounter challenges in differentiating relevant data features from irrelevant data features (Canham & Hegarty, 2010; Jeong et al., 2007; Kanari & Millar, 2004; Shah & Hoeffner, 2002), which may cause individuals to identify less relevant patterns in the data (Zagallo et al., 2016). Once the individual identifies the different patterns in the data, they must connect the pattern to the underlying scientific phenomenon to form a claim about the data (Carpenter & Shah, 1998; Glazer, 2011; Latour, 1999; Shah & Hoeffner, 2002). Essentially, the last step serves as the opportunity to make meaning of the data by using one's knowledge of the underlying content (Carpenter & Shah, 1998; Roth & Bowen, 2000; Shah & Hoeffner, 2002). Not fully engaging and completing this last step can result in more superficial interpretations of the data (Lai et al., 2016).

Prior Knowledge in Data Analysis

There is a growing body of evidence that suggests that what prior knowledge is accessed during data and graph analysis greatly affects what meaning one can make from a data set. Graphical representations often elicit one's schemas of graphs to help orient oneself to the information presented and make sense of it (Pinker & Feedle, 1990; Shah & Carpenter, 1995; Shah & Hoeffner, 2002). For instance, it is a common graphical convention to place independent variables on the x-axis of the graph and dependent variables on the y-axis of the graph. When this convention is not followed, interpreters often make faulty interpretations due to their schema of where independent and dependent variables exist on the graph (Shah & Hoeffner, 2002). One's knowledge of the content in the data representation or data set also greatly affects how one engages in data analysis. For instance, having prior knowledge of the content in the graph can help analyzers encode more relevant information, identify relevant patterns in the data, and engage in a more thorough analysis of the data set (Canham & Hegarty, 2010; Roth & Bowen, 2000; Shah & Hoeffner, 2002). Conversely, not accessing the necessary content knowledge can affect what information is deemed important and consequently impact what patterns are identified (Carpenter & Shah, 1998; Jeong et al., 2007; Masnick & Morris, 2022; Shah & Hoeffner, 2002).

Research in chemistry education also highlights the importance of prior knowledge in the students' data analyses, in that multiple studies have identified that undergraduate students rely on heuristics and "everyday" knowledge when reasoning with chemical data (Becker et al., 2017; Heisterkamp & Talanquer, 2015; Zhou & Moon, 2023). In one study, Becker and colleagues identified the different ways that undergraduate students used chemical rate data to construct mathematical kinetic models (2017). Some students relied on the stoichiometric constants from the balanced chemical reaction equations rather than consulting the initial reaction rate and concentration data. They also used heuristics that were not productive in constructing an empirically sound kinetic model. It is possible that this could be due to students not accessing relevant content knowledge related to reaction rates, and instead relying on other knowledge that they could access to aid in making sense of the data. Similarly, Zhou and Moon's study found that when asked to decide on which reaction mechanism was the most likely, some undergraduate students admitted to not using any of the data presented to them to evaluate (2023). Instead, students relied on heuristics to determine the mechanistic pathway, such as reasoning that one of the pathways seemed to use less energy, even though none of the data presented related to energy. In Heisterkamp and Talanquer's case study of an undergraduate chemistry student analyzing boiling point data and ionization energy, the authors found that the participant often used their everyday knowledge in tandem with their content knowledge (2015). For example, the participant used their knowledge of the density of macroscopic items like rocks and cotton with their knowledge of electron density when explaining ionization energy trends. From these studies, it is evident that what prior knowledge students can access when analyzing data affects how they engage in the analysis of chemical data.

Although much of the research in chemistry education demonstrates the importance of accessing prior knowledge, it is also important to consider how the student uses the prior knowledge in their data analysis. For instance, Connor and colleagues identified that both undergraduate and graduate chemists use the "n+1" rule when analyzing proton nuclear magnetic resonance (^{1}H NMR) spectra to determine molecular structure, but graduate students knew under what contexts this rule was no longer valid (2021). Additionally, Zhou and Moon's study found that undergraduate students who used data to form their conclusions and students who did not use the data often both employed the same content knowledge related to the rotation of bonds in a cyclic molecule (2023). However, those who used the data could identify the connections between the data provided and the knowledge that they were accessing to construct better-supported conclusions. Both studies suggest that simply activating necessary prior knowledge can only take one's data analysis so far.

More research is needed to uncover how prior knowledge is used in the data analysis process. Specifically, there is a need to establish a model that can describe the underlying mechanism of how prior knowledge is used throughout the data analysis process. This can best be fulfilled through qualitative inquiry, in which think-aloud interviews are used to elicit when and how students draw upon their prior knowledge and experiences to make sense of data. These interviews can help provide a rich, detailed source of information directly sourced from students to provide a fuller picture of the phenomenon (Creswell & Poth, 2016; Merriam & Tisdell, 2016). Additionally, qualitative analysis of interviews can best capture the complex and sometimes even messy processes students engage in as they make sense of data (Creswell & Poth, 2016; Merriam & Tisdell, 2016). In this way, a theory can be built that captures a mechanism of how prior knowledge is used throughout students' data analyses (Creswell & Poth, 2016; Merriam & Tisdell, 2016).

Epistemic Criteria Used in Data Analysis

When engaging in data analysis, all students must consider several epistemic criteria for what constitutes a quality data analysis. These epistemic criteria help guide students in their reasoning to develop more sophisticated and complete analyses. A deep understanding of the epistemic criteria related to data analysis (and science practices more generally) is demonstrative of further developing competency in the domain of data analysis (Kuhn et al., 2017). There are several criteria for students to consider including (but not limited to) the following:

Students must consult and use data as evidence when forming a conclusion (National Research Council, 2012). It is only in this way that scientific knowledge is constructed; empirical data must serve as the basis for scientific conclusions (Duschl, 2008; Ford, 2012; Jiménez-Aleixandre & Crujeiras, 2017; Kuhn et al., 2017; Taber, 2017). When students base their analysis on heuristics or intuition alone, they risk forming poorly supported conclusions or misinformed conclusions that are not scientific in nature (Becker et al., 2017; Zhou & Moon, 2023). Thus, a strong data analysis is built upon and supported by the foundation of empirical data.

Students must also consider what knowledge to use to make sense of the data and consider if that knowledge is most appropriate for their analysis (Duncan et al., 2018; Grolemund & Wickham, 2014; Klein et al., 2007). This is important as students may use “everyday” knowledge in place of more relevant prior knowledge when analyzing data until prompted otherwise (Heisterkamp & Talanquer, 2015). Students may also draw upon conceptual knowledge that is related to the data but is not productive for the context. For instance, Zhou and Moon found that when undergraduate students analyzed data on halogenated products to identify a mechanism that produced the products, students used their knowledge of electronegativity to make sense of data (2023). Although the data set contained data on halogens with different electronegativities, the mechanisms presented did not depend on the electronegativities of the halogens. Had the undergraduate students considered other possible knowledge and interrogated which could best make sense of the data as part of their analysis process, they might have identified that electronegativity was not productive in identifying the mechanism that produced the halogenated product yields. Therefore, not only do stronger analyses use conceptual knowledge, but stronger analyses use conceptual knowledge that offers the most explanatory power for making sense of the data.

Additionally, students must recognize when there is discrepant data, and then consider how they must engage with said discrepancy (Grolemund & Wickham, 2014). When making sense of data, the data should serve as evidence to support one’s conclusions formed from one’s analysis (National Research Council, 2012). Thus, when students encounter discrepant data, they must decide if the data weakens or limits their conclusions, and if so, what they must do to address said discrepancy.

To develop and internalize epistemic criteria such as these, it is necessary to implement more opportunities for students to practice them in the classroom. However, it can be difficult to attend to individual students' engagement in data analysis and interpretation in the classroom, as instructors and teaching assistants must monitor and manage many groups of students in most classroom spaces (Phillips et al., 2021). Although some work has investigated methods to implement data analysis in the undergraduate biology classroom (Bolger et al., 2021; Zagallo et al., 2016), there is still a need to investigate more classroom methods that can develop this competency, especially within the context of the chemistry.

Peer Review

Many of the science practices, including data analysis, will produce written knowledge products. Peer review can serve as a tool to evaluate these written products. By being exposed to other students' products when giving feedback and receiving compelling feedback from others, students can develop and practice evaluative judgment (Nicol et al., 2014; Sadler, 2010). This evaluative judgment serves as a "rubric" of what constitutes a successful and quality product for a specific domain (Sadler, 2010). Developing and exercising evaluative judgment related to science practices, like data analysis, is equivalent to building a deep understanding of the epistemic criteria of science practices, which is essential for developing competency (Kuhn et al., 2017).

Peer review serves as a vehicle to produce different sources of feedback on written work. Receiving feedback from multiple peers in a peer review setting offers students the opportunity to improve their work without needing the same feedback from an instructor (Cho & MacArthur, 2011). However, students must develop feedback literacy to effectively engage with feedback that they receive from peers. This involves students appreciating the feedback and recognizing its value, exercising their evaluative judgment on the feedback (i.e., does the feedback help students better fulfill the criteria for success), managing affect relating to the received feedback (e.g., managing defensive feelings, doubt, uncertainty, etc.), and then acting on the feedback that they have received (Carless & Boud, 2018). This can be a laborious literacy to develop and purposefully engage in, especially in the context of large-enrollment courses where students may not know the peers whom they receive feedback from. Given the laborious nature of these steps, it is not surprising that receiving feedback in peer review may not compel some students to make necessary changes to their work (Finkenstaedt-Quinn et al., 2021).

Some studies have shown that students make more improvements to their written work after giving feedback in peer review (Anker-Hansen & Andrée, 2015; Cho & MacArthur, 2011; Ion et al., 2019; Lundstrom & Baker, 2009; Nicol et al., 2014; Nicol & McCallum, 2021). Students sometimes report that giving feedback helped them identify areas of improvement in their work before they had even engaged with the feedback that they had received from peers (Nicol et al., 2014). Considering that students demonstrate such significant improvements from giving feedback, it is necessary to uncover the underlying mechanism that produces these improvements.

When students give feedback in peer review, they begin by making a comparison between their own work and the work that they are reviewing (McConlogue, 2015; Nicol, 2020; Nicol et al., 2014; Nicol & McCallum, 2021; van Popta et al., 2017). During this comparison, students use their own work as a reference. Engaging in this comparison between the two products can prompt students to reflect on the task's criteria for success and exercise evaluative judgment on their own work (McConlogue, 2015; Nicol et al., 2014; Nicol & McCallum, 2021). This process can help students generate internal feedback which can be used to inform and improve their work to better meet the criteria of the task (Anker-Hansen & Andrée, 2015; Nicol, 2020; Nicol et al., 2014; Nicol & Kushwah, 2023; Nicol & McCallum, 2021). Given the research on giving feedback in peer review, questions remain on how to effectively implement peer review within chemistry classrooms.

Recent studies within chemistry education have attempted to describe the role of peer review in undergraduate students' revisions in writing assignments (Finkenstaedt-Quinn et al., 2024; Watts et al., 2024); however, their decision to only consider undergraduate students' written peer review comments and final written products limits the findings that they produce. In both studies, the authors identified new writing present in a student's final written draft that seemed to reflect feedback that the student had given or received. Though the authors could infer that the peer review's feedback caused revisions, in no part of the study were students given the opportunity to explain why they chose to revise or not. Additionally, students were not asked how the different sources of feedback influenced their decision-making to revise. This makes it difficult for instructors to know how to best design classroom peer review assignments to elicit students' full participation to produce their best-written work.

Thus, research is needed to develop a theory that can explain how and why students choose to revise their written work when giving feedback, especially in the context of science practices. To do this, qualitative interviews are needed so that students can directly explain how they evaluated their written work after giving feedback to others (Creswell & Poth, 2016). These interviews can be analyzed via exploratory qualitative inquiry to describe the rather complex feedback generation and implementation processes found in peer review (Creswell & Poth, 2016;

Merriam & Tisdell, 2016). In this way, a more comprehensive model of giving feedback in peer review can be built to provide more explicit implications to researchers and instructors (Creswell & Poth, 2016; Merriam & Tisdell, 2016).

Graduate Students' Sensemaking of Unexpected Data

One of the central goals of chemistry graduate programs is to develop competent and independent scholars within the field of chemistry (National Academies of Sciences, 2018; The American Chemical Society, 2012). To do this, chemistry departments require that graduate students participate in and fulfill different programmatic elements, such as writing and defending a dissertation (Donkor & Harshman, 2023). Working towards these programmatic elements requires that graduate students analyze empirical data and conduct original research that contributes new knowledge to their field (Donkor & Harshman, 2023).

Scientists often construct new knowledge for the world through engaging with uncertainty (Kampourakis & McCain, 2020). A form of uncertainty that scientists, including chemistry graduate students, face in their research is data that is discrepant with one's expectations (Grolemund & Wickham, 2014). For instance, scientists stage experiments with uncertain outcomes and make sense of the results as part of their research (Latour, 1999). Thus, to develop into competent and independent scholars who contribute new knowledge to their field, chemistry graduate students must be trained to productively engage with unexpected data.

Few studies in chemistry education have investigated chemistry graduate students' reasoning with chemical data, and the ones that have done so primarily focus on graduate students' sensemaking of IR and ^{1}H NMR spectra to determine chemical structures. In a study of how chemistry graduate students used IR and ^{1}H NMR to construct a molecular structure, Cartrette and Bodner qualitatively captured the different ways in which more successful students differed from less successful students (2010). The authors found that more successful graduate students searched the data for less salient data, specifically the ^{1}H NMR spectra's coupling constants, to aid them in determining the structure of the molecule. Furthermore, more successful graduate students were more likely to check their proposed structure against the spectra, which suggests that students were verifying that their claims were supported by the evidence provided. In another study, Connor and colleagues used eye tracking and retrospective think-aloud interviews to explore undergraduate and graduate students' analyses of IR and ^{1}H NMR to determine if a synthesis was successful or not (2021). The authors identified that graduate students spent less time than undergraduates looking at the fingerprint region of the IR spectra. The authors argue this reflected that the graduate students knew the fingerprint region has varying useful information for determining molecular structure. Additionally, the graduate students' gaze patterns indicate that they searched for complementary data between the IR and ^{1}H NMR data to help them determine if a synthesis was successful or not. This could be indicative of graduate students looking for multiple sources of evidence to support their claims.

Both studies from Cartrette and Bodner and Connor *et al.* have documented ways in which chemistry graduate students make sense of IR and ^{1}H NMR spectra to determine molecular structures, but there are still gaps on graduate students' sensemaking of data overall. Specifically, neither study considers how graduate students respond to data that they do not expect. Given that chemistry graduate students are likely to engage with unexpected data as they conduct their research and that many graduate students face challenges in their development as independent scholars (Gardner, 2007, 2008; Grolemund & Wickham, 2014), there is a need for research to investigate graduate students' sensemaking of discrepant data. Specifically, research is needed that can qualitatively document how chemistry graduate students respond to data they do not expect. This can best be achieved by interviewing graduate students as they reason through unexpected data to provide detailed accounts of their sensemaking (Merriam & Tisdell, 2016). Directly interviewing graduate students as they make sense of unexpected data can provide a more complex, nuanced account of their reasoning (Creswell & Poth, 2016). In this way, qualitative inquiry can help to develop a model that describes the different ways that chemistry graduate students respond to discrepant data (Creswell & Poth, 2016; Merriam & Tisdell, 2016). This model can then be used to provide suggestions that help guide students' analyses of unexpected data.

Study Goals and Research Questions

This dissertation seeks to advance knowledge of undergraduate and graduate chemistry students' data analysis in three different ways. All three studies are qualitative and use student interviews to build models that describe various processes undergraduate and graduate students use while reasoning through different data analysis tasks. The studies are summarized below.

The first study investigated how undergraduate chemistry students' prior knowledge and experiences were activated and leveraged as students completed a data analysis task. Undergraduate chemistry students participated in think-aloud interviews as they analyzed a line graph to identify ideal experimental conditions for a dual-phase extraction and constructed an argument for their choice using the experimental data. The following research question guided the thematic analysis of this study:

1. How do general chemistry students' prior knowledge and experiences interact with their graph analysis during data analysis and interpretation?

The next study explored how a simulated peer review could be used to prompt undergraduate students to engage in self-evaluation of their written knowledge product related to a science practice, specifically data analysis. To do this, undergraduate chemistry students were interviewed as they compared their responses from a data analysis task to three preconstructed sample responses that each had an embedded flaw. Each comparison constituted a "social comparison" in which students engaged in a critique of both the sample response and their own response. The interviews were analyzed using social comparison and internal feedback theories to construct a model of students' evaluations of their written work. The research question that guided the analysis of the study was:

2. How do chemistry students evaluate their own data interpretations when critiquing hypothetical peers' data interpretations?

The final study characterizes how chemistry graduate students responded to and made sense of empirical data that did not align with their expectations in some way. Chemistry graduate students participated in a think-aloud interview as they analyzed multiple data sets to construct an explanation for the cause of a chemical phenomenon. The discrepant moments arose naturally as students voiced that some aspects of the data did not align with their expectations. The study used Data-Frame Theory to characterize how chemistry graduate students responded to their discrepancies, and in what ways their responses to the discrepant data impacted their sensemaking. The research questions that guided the study were:

3. How do chemistry graduate students respond to discrepant data when analyzing multiple data sets to explain a chemical phenomenon?
4. How can different responses to discrepant data differentially affect sensemaking?

Overview of Chapters

The following dissertation is formatted as a paper collection. The second chapter describes the methods, analysis, and results of the first study; the third chapter describes the methods, analysis, and results of the second study; and the fourth chapter describes the methods, analysis, and results of the third study. The fifth chapter will discuss conclusions, implications for both research and practice, and recommendations for cultivating students’ data analyses and implementing peer review in the classroom.

CHAPTER 2
A CHARACTERIZATION OF CHEMISTRY LEARNERS' ENGAGEMENT IN DATA ANALYSIS AND INTERPRETATION

Abstract

Both graph comprehension and data analysis and interpretation are influenced by one's prior knowledge and experiences. To understand how one's prior knowledge and experiences interact with their analysis of a graph, I conducted think-aloud interviews with general chemistry students as they interpreted a graph to determine optimal conditions for an experiment. Afterward, students engaged in a simulated peer review by reviewing three sample responses, which further revealed their reasoning. I deconstructed students' analyses using Data-Frame Theory to identify the prior knowledge and experiences that informed and guided their analysis, as well as characterized moments in which their analysis was influenced by different sources of information. Using template analysis, I present and discuss four themes: establishing the frame, observing and interacting with the data, data-frame interactions, and when frames change. From these findings, I discuss the implications of utilizing students' prior knowledge and experiences to aid in their data analysis and interpretation, as well as identify opportunities for future research.

Introduction

Reforms in science and chemistry education have emphasized the need for STEM students to engage in science and engineering practices (Cooper & Klymkowsky, 2013; National Research Council, 2012; Talanquer & Pollard, 2010). At the K-12 level, the Next Generation Science Standards in the United States have effectively outlined competency across eight science practices for various grade bands (2012). While it is critical for undergraduate students to develop competency in these practices to be scientifically literate and prepared for future careers in STEM (Cooper et al., 2015), research is required to define competency and outline how it develops. As part of this research, I aim to characterize how undergraduate chemistry students engage in the science practice of data analysis and interpretation.

Studies across science education, educational psychology, and chemistry education have shown that prior knowledge and experiences influence students' data-based reasoning. Accessing (or not accessing) certain prior knowledge and experiences can affect the features and patterns that one notices in a data set (Friel et al., 2001; Heisterkamp & Talanquer, 2015; Jeong et al., 2007; Pinker & Feedle, 1990; Shah & Carpenter, 1995; Shah & Hoeffner, 2002). Relevant prior knowledge is also necessary to tie patterns found within the data back to the phenomenon under study (Lai et al., 2016; Latour, 1999; Shah & Carpenter, 1995; Shah & Hoeffner, 2002). In the absence of the necessary content knowledge, many students will rely on heuristics and intuition or neglect to use reasoning entirely (Becker et al., 2017; Heisterkamp & Talanquer, 2015; Masnick & Morris, 2022). For more advanced scientists, this prior knowledge is key because with it, the scientist will contextualize their interpretations in the broader scientific context; that is, they consider hypotheses, theory, experimental design, and implications to draw conclusions (Angra & Gardner, 2017). To support the development of this integration between relevant prior knowledge and science practice engagement for younger scientists, there is a need to understand how these two domains interact.

To that end, this study seeks to investigate how general chemistry students use data to determine optimal conditions for an experiment. I use Data-Frame Theory to model the dynamic interactions between a student's prior knowledge and experiences and their analysis of a graph as they engage in a data analysis and interpretation task. The study is specifically guided by this research question: How do general chemistry students' prior knowledge and experiences interact with their graph analysis during data analysis and interpretation?

Background

Data Analysis and Interpretation

When encountering data, an individual must first encode the visual features and identify the features that are important (Glazer, 2011; Shah & Carpenter, 1995; Shah & Hoeffner, 2002; Zagallo et al., 2016). Once the important visual features are encoded, any relevant patterns or relationships must be identified within the data displayed (Ratwani et al., 2008; Zagallo et al., 2016). Not being able to differentiate between the relevant and irrelevant information has been noted as one of the many challenges for data-based reasoning (Jeong et al., 2007; Kanari & Millar, 2004). In addition to this, students may use select data to form conclusions or predictions, often explaining away the data that contradicts their theory (Chinn & Brewer, 1993, 2001; Meister et al., 2021). After the important patterns have been identified, they must be connected back to the phenomena or concepts modeled in the representation (Glazer, 2011; Latour, 1999; Shah & Hoeffner, 2002). Both identifying relevant information and connecting the information back to phenomena are processes that are largely influenced by an individual's prior knowledge and experiences with the type of data representation and the phenomenon being considered. Graph schemas aid in identifying the kinds of relationships represented in the graph (Pinker & Feedle, 1990). Prior knowledge of the phenomenon modeled in the graph can also "unlock" more expert-like reasoning and comprehension (Roth & Bowen, 2000). Not having ready access to the content knowledge of the phenomenon could affect what visual features and patterns are identified within the graph (Shah & Hoeffner, 2002). This is also true of analysis with other forms of empirical data in representations such as data tables (Jeong et al., 2007; Masnick & Morris, 2022).

Many of the challenges identified for students' data-based reasoning in chemistry education echo those of science education and psychology. In a study investigating students' use of initial kinetic rates data in constructing rate laws, Becker and colleagues found that some students neglected to use some of the empirical data provided or even neglected to use the data entirely (Becker et al., 2017). They also found that many students used unproductive reasoning and heuristics when forming their models. I posit that the use of this reasoning could potentially come from students' lack of relevant content knowledge of the phenomena they were modeling. In another data analysis and interpretation study, Heisterkamp and Talanquer used a case study approach to explore a participant's analysis of boiling point data and ionization energies (2015). The authors identified the use of "hybridized" reasoning wherein the participant used a mix of intuitive ideas and chemical content knowledge to form explanations for trends in the data. The participant also relied on explicit surface features of data to form explanations. For example, the participant tried to explain the increasing boiling points of substances by counting the number of atoms in each molecule and calculating the differences in the masses of compounds. In both studies, chemistry students' reasoning seems to be influenced by their prior knowledge and experiences that they use to analyze the data. Additional work in biology education and science education has further supported this (Angra & Gardner, 2017; Jeong et al., 2007).

The work in graph comprehension, data analysis and interpretation, and chemistry education has continually found that prior knowledge and experiences affect students' analyses with data representations (Carpenter & Shah, 1998; Glazer, 2011; Heisterkamp & Talanquer, 2015; Jeong et al., 2007; Masnick & Morris, 2022; Pinker & Feedle, 1990; Roth & Bowen, 2000;

Shah & Carpenter, 1995; Shah & Hoeffner, 2002). However, little work has considered how prior knowledge and experiences interact with one's data-based reasoning. Therefore, I sought to bridge this gap in the literature by characterizing the processes by which chemistry students engaged in data analysis and interpretation using real data. I specifically aimed to account for their prior knowledge and experiences that they used to make sense of the data.

Data-Frame Theory

The science practice of data analysis and interpretation can be viewed as a sensemaking process (Chen and Terada, 2021). Raw empirical data must be manipulated, organized, and interpreted by the scientist to generate meaning (National Research Council, 2012). One kind of data representation that all scientists encounter, regardless of what field they are in, is the graph. Numerous studies have characterized the graph comprehension process (Carpenter & Shah, 1998; Friel et al., 2001; Ratwani et al., 2008; Shah & Carpenter, 1995), but it is typically detached from science practice. These studies offer little insight into how an analyst's prior science knowledge and science practice competency interact when they analyze and interpret data. With the specific aim of understanding this interaction, I use the Data-Frame Theory of sensemaking (Klein et al., 2007; Klein & Moon, 2006).

Data-Frame Theory asserts that analysts concurrently construct data (observations) and frame (reasoning), and that the data and frame inform one another. The frame is an explanatory structure that helps to account for pieces of data by describing their relationship to other data in the environment. In this way, the frame serves as a lens for making observations and assigning meaning to those observations. Frames can take the structures of stories, scripts, maps, or plans, and are synonymous with schemas (Gouvea et al., 2019; Hammer, 2000; Klein et al., 2007).

The other component of data-frame theory is data. Data is the information extracted from the environment. Within chemistry, data would be considered the empirical data collected from experiments, but other sources of information, such as experimental schemas, directions, molecular structures, or events, can be considered data as well.

Data and frames tend to interact in a cyclic pattern. A person begins by encountering data in an environment, and certain features of the data can cue a frame. Certain points of data or key features within the data serve as anchors that help to elicit a frame. This frame can then guide how the analyst makes sense of the data. The frame is influenced by the analyst's prior knowledge and experience with similar data. The frame can also be influenced by whatever goals the person may have associated with the data (e.g., a hypothesis). Once the frame is established, a person can begin to search for more information. During this search the frame "filters" incoming information to help a person seek more relevant information to aid in sensemaking. To account for the incoming information, the frame can be elaborated and extended to fill in whatever gaps were not originally accounted for, as illustrated in Figure 2.1.

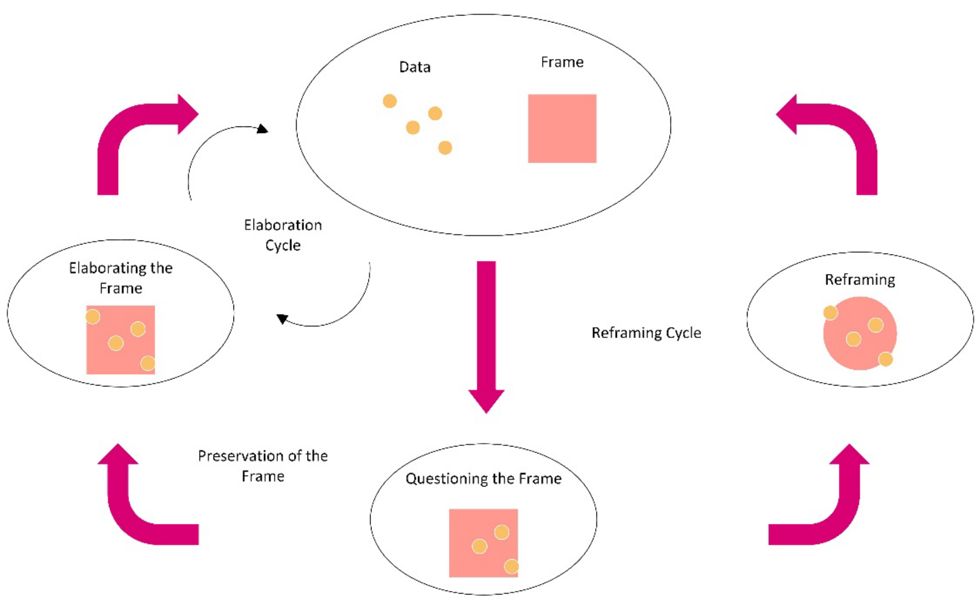

Figure 2.1: Data-Frame Theory Model of Sensemaking modified from Klein and Moon (2006)

During the sensemaking process, it is likely that an individual will encounter data that is inconsistent with their frame. When this happens one can begin to question their frame and note where the data violates expectations generated from the frame. After these gaps are exposed, the frame can be maintained through two processes: preserving the frame or elaborating the frame. To preserve the frame, the anomalous data can be explained away or disregarded. This aligns with many other studies within science education and discipline-based education research in that many will discount data that contradicts their mental model (Chinn & Brewer, 1993, 2001; Meister et al., 2021). If the contradictory data is accepted, one can engage in elaborating their frame. During this process, the frame is expanded and extended to account for the new information. Although the frame might be undergoing changes, the frame's integrity is still intact, and its key anchors are maintained.

If the anomalous data is accepted and cannot be integrated into the frame, a person may decide to disengage from their current frame and construct a new frame. In this reframing process, new anchors will be searched for to establish the new frame. Data that was previously overlooked or discarded may now be considered and interpreted from a new perspective as well.

At times an individual might compare or even consider multiple frames to make sense of the data. Klein and colleagues estimate that a person can use up to three frames simultaneously (Klein et al., 2007). Using multiple frames in sensemaking is akin to observing the data from multiple perspectives. Each frame will have different anchors that are unique to each frame, allowing the individual to pick up on different aspects of the data specific to the perspective they are considering. These frames likely have different key anchors unique to each frame allowing the individual to pick up on different pieces of data through each perspective they are considering. Students can also compare the frame that they are using to another frame they encounter. This can potentially cause students to disengage from their initial frame and use the other if they find the new frame compelling enough.

I posit that for an individual to consider and compare a frame against their own, they likely need to engage in decentering from a central perspective. Decentering involves recognizing alternative perceptions and reasoning for the same problem or situation (Piaget, 1955). To consider and compare frames, an individual must recognize alternative perspectives of the same data, which means they are recognizing alternative perceptions of the same problem. Differentiating between different perspectives in this way has been shown to support more productive argumentation (Moon et al., 2017) and support teacher noticing (Teuscher et al., 2016).

Data-Frame Theory also reports that experts and novices engage in data-based reasoning in similar ways. The primary difference between an expert's analysis and a novice's analysis is what mental models and prior knowledge the individual can access. Experts possess richer collections of conceptual knowledge and more experiences that can inform their frame when engaging in sensemaking. In having richer collections of knowledge and experiences, experts can access more frames to use when engaging in sensemaking. This also allows experts to find deeper meaning in data that is presented to them, whereas novices tend to produce more shallow conclusions. Data-Frame Theory helps to identify what pieces of knowledge and experiences inform an expert's frame to lead to more productive sensemaking and sophisticated conclusions.

I propose using Klein's Data-Frame Theory primarily because it enables us to deconstruct an analysis, as well as document interactions between frames and data. Other studies have shown the utility of using a frame theory to explore how one's frame can affect the ways in which one reasons with a problem (Hammer et al., 2004; Slominski et al., 2020). Similarly, Data-Frame Theory can help to identify a student's different prior knowledge and experiences they use to guide their analysis of a dataset. Different data features can also be identified that were used to help students reach a conclusion after consulting the dataset. Data-Frame Theory also can characterize more complex interactions between data and one's frame. Inconsistent and anomalous data or information that challenge a scientist's initial ideas and hypotheses can be explored and characterized. Additionally, data analysis and interpretation, like other science practices, is an iterative and dynamic process. Scientists' previous experiences with types of data and knowledge of the content they are studying greatly influence how they interact with their own data. Much of their data analysis process is informed and guided by the scientists' expertise. When they encounter anomalous data, scientists must consult their prior knowledge and decide how they plan to address it. In this work, I am particularly interested in accounting for these interactions to characterize the different processes that a chemistry student engages in while using data from a graph to reach a conclusion.

Data Collection and Analysis

Interview Protocol

The semi-structured interviews used for this study consisted of two stages. In the first stage, students determined the optimal reaction conditions for an experiment using a graph. The task is a scaffolded version of the same experimental decision made by Doidge and colleagues when deciding what concentration of HCl would best be used to isolate gold in a dual-phase extraction of waste electronic equipment (Doidge et al., 2016). Students were specifically tasked to use the graph from Doidge and colleagues to choose what concentration of HCl would best extract a maximal amount of gold and a minimal amount of tin and iron, as shown in Figure 2.1.

They chose between 0 M, 2 M, and 4 M. Throughout this part of the interview, students were encouraged to annotate the graph using the Zoom annotation feature to help explain their reasoning, especially when referring to specific parts of the graph. At the end of this interview phase, students wrote a summary of their verbal analysis and were told to include whatever amount of detail and evidence they deemed necessary to convince someone else. The written summaries were typed into the chat function of the Zoom call to be referenced later in the interview.

In the second stage of the interview, students engaged in a simulated peer review, in which they gave feedback to three sample responses and compared them to their own (Berg & Moon, 2022). Each sample response was pre-constructed to support one of the three choices for the task. The responses also had different argumentative flaws embedded for students to identify and critique. During the simulated peer review, students compared their own analysis to that of the sample response and gave feedback meant to improve the response. This comparison helped to indirectly probe the student's own analysis as well, especially if they mentioned that the analysis of a sample response was similar to their own. During the simulated peer review, several students expressed feeling less confident in their selection of which concentration of HCl to use for the task. To remedy this, students were given the opportunity to make changes to their work. If students made these changes during their interview, they were asked how their response had changed from before the peer review.

See Appendix A for the full interview protocol.

Sampling

This study took place in a large, Midwestern university in the fall of 2020. Institutional IRB approval was obtained before recruiting for interviews. Participating students (N=18) were from general chemistry I and II courses for both majors and non-majors. Students submitted consent forms electronically before their interviews. After the interview, students were compensated with a $20 gift card for their participation.

Data Collection

All interviews were conducted remotely via Zoom. Each interview was recorded within the application, and the recordings were used to generate transcripts of the interview. The video recordings of the interview were kept and used in the analysis for visual reference, especially when students annotated the graph in reference to something. All transcripts and videos have been deidentified and pseudonyms have been given to participants. Any stills from the video recording consist of only the graph and the students' drawing; any references to the students' names have been cropped out of the image when used in analysis.

Analysis

I analyzed transcripts from both stages of the interview for analysis. Particular attention was paid to the second stage if students changed their answer to a different concentration of HCl. I believe that this was evidence of students engaging in the task with a new or elaborated frame.

Because my research question focuses on the processes that students are engaged in when analyzing a graph, I used both process coding and open coding (Miles et al., 2014). Process codes helped to identify different actions that the students took at different times during the task. The codes were developed to describe actions that the students consciously described they were doing as well as actions that they may not have been aware of. The open codes that were developed at this time were used to capture different features of the graph that students were using as well as the different conceptual and everyday ideas that they used to help with the task.

After open coding, I employed template analysis, a type of thematic analysis, to organize codes (Brooks et al., 2015). I began by sorting open codes into two a priori categories: the data category and the frame category. Codes that had anything to do with information that could be found in the graph were sorted into the data category. These codes included instances in which students pointed out peaks within the graph or comparisons students may have made between different parts of the graph. Codes that focused on reasoning with information external to the graph or used information from the prompt were sorted into the frame category. Many of these codes included some sort of set of goals or objectives that students mentioned that they wanted to accomplish to fulfill the task. These are assumed to be parts of a student's frame that helped guide their analysis.

In the second round of template analysis, I noticed that some of the codes sorted into the data category also had elements of a frame to them as well. These codes did have elements of a graph feature included in them, but there was also some sort of opinion or evaluation being made with the data. For example, in his comparison between the increases of gold and other metals between 2 M and 4 M, Fernald said, *"And this increase [in other metals] is not compensated for the increase in gold accumulated."* This part of his analysis does include directly observable information from the graph (the increases), but there is also an element of value or meaning ascribed in the comparison that is not directly observable (the compensation of one increase over the other). I inferred that assertions and evaluations like Fernald's showcase an interaction between the student's frame and the data; thus, a new coding category was constructed to capture them.

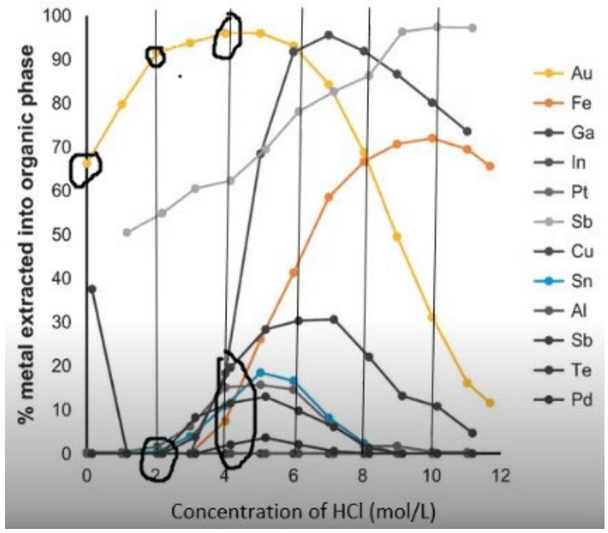

Figure 2.2: Screen capture of Fernald's graph

After using template analysis to organize a student's analysis, I developed a coding scheme in which the analysis is followed and deconstructed according to three categories: the frame, the data, and the data-frame interactions. To identify the student's frame, I identified key anchors that seemed to guide their analysis. I identified these anchors by considering three things: how a student, explicitly or implicitly, defined minimal or maximal for the task; what concentrations of HCl were being considered (i.e., if a student constrained their focus to one area of the graph or used data from beyond the designated choices to inform their choice); and what metals were being considered. I also tried to identify any outside knowledge, conceptual or previous experiences that seemed to inform or guide parts of the frame. Next, I identified the data being used by the student during the think-aloud interview. This meant identifying which points or areas of the graph the student deemed important, identifying what comparisons students made in the graph, and how such comparisons were made. Finally, I identified the data-frame interactions that occurred during the student's analysis. These interactions were moments in which the frame was used to inform decisions with the data and moments in which the data had some sort of effect on the student's frame. Each student's interview was deconstructed following this method of analysis.

I recognize that in some interviews, students' analyses were not as linear or direct as others. Many students changed their answers at some point during the interview. The change in answer likely indicated that something had changed in the student's frame. To distinguish between a student's different frames, I identified the anchors that seemed to inform the student's answer before and after the change. For this part of the analysis, I identified either an entirely new set of anchors for a student's frame or I identified new anchors that might have been added to the pre-existing set. After this, I identified the circumstances under which the frames changed as well as pinpointed the actions students undertook when their analysis deviated or changed somehow.

Trustworthiness

An outside researcher who is also trained on Data-Frame Theory as a theoretical framework was recruited to establish trustworthiness for the modeling of students' data analyses. The outside researcher and I collaboratively deconstructed three students' interviews to identify the students' frames for analysis, the data used, and the data-frame interactions that took place to help a student reach their final answer. Any disagreements or discrepancies between researchers were discussed and changes were made to the coding to reflect the discussion.

To begin, the outside researcher was trained on the anchors that I had identified as relevant to the task. The outside researcher and I then constructed frames for each student by identifying anchors that seemed to guide the student through their analysis. There was an initial discrepancy concerning whether one student had formed a frame and immediately changed to another frame or if the student had been simultaneously using two frames to approach the task before settling on one. The outside researcher and I discussed and decided that because the student seemed to weigh both frames equally when speculating, he likely was considering multiple frames at once rather than changing frames. This was noted and applied to similar interviews in which students considered multiple answers until eventually settling on one. The researchers also worked to describe new frames that students had after changing answers in the interview.

Once frames had been constructed for each student, the outside researcher and I identified data that each student considered in their analysis, as well as what students did with the data. The interview transcripts were read line by line to identify the graph features students spoke about or marked on their screens. We then identified what students did with the graph features to help further their analysis.

After considering the frames and the data that students used within their analysis, the outside researcher and I identified data-frame interactions for each interview. Much of the discussion focused on two of the interviews in which students changed their answers at some point. We inferred that a change in answer meant that students experienced some sort of change in their frame. We further inferred that there was data to prompt the change. Both students were exposed to new information before they changed their answer: one student reread the prompt that changed their perspective and another student engaged in a peer review that changed their perspective. Therefore, in the instance of a frame change, we decided to classify it as a data-frame interaction, and to expand the definition of data to include different sources of information such as simulated peer review and rereading of a prompt. Following this, other interviews in which students experienced a change in frame were reanalyzed to identify the information that caused such a change.

Once consensus was reached for all three interviews, I coded the remaining interviews.

Results

Overview

The first theme examined is the forming of a frame. Chemistry students began the task by establishing some sort of frame to help make sense of the data. This often involved first drawing upon relevant prior knowledge and experiences that were activated by the task. This in turn would help students establish goals and objectives to guide their analysis and completion of the task. Although all students did establish a frame to work through the task, many of the students began the task using multiple frames in their approach before deciding on one.

The next theme investigated is the theme of observing and interacting with data. Having established and decided on a frame, students then began searching for graph data that was relevant to their frame. This information included different points within the graph such as the peaks for different lines and low points for others. Once students had identified the relevant information, they would engage in interacting with the graphical data. For this task, these interactions mainly consisted of comparisons of different graphical data.

Following the themes of frame and data, a third theme is introduced that examines the interactions between the two. After exploring the data to find relevant information and engaging with important graphical data, students engaged in various data-frame interactions. The most common data-frame interaction involved students using graphical data to make evaluations of the different HCl concentration options. These evaluations, in turn, were used to help students gauge which option best fulfilled the task according to their frame. For some students, the data within the graph activated another data-frame interaction in which students deviated from the task's prompt and incorporated extra or different objectives in their frame that were not observed in other students' analyses.

The final theme explored is when students' frames changed during their analysis. I consider this to be another form of a data-frame interaction; however, the interaction resulted from incoming information outside of the data in the graph. One source of information that prompted such a change was the rereading of the prompt. Students began their interviews with one answer informed by one frame, but immediately after rereading the directions changed their answer informed by a different frame. Another type of incoming information that led to many students changing their answers was considering an alternative perspective. Typically, this occurred during the simulated peer review in which students were exposed to an alternative frame that seemed to prompt a change in perspective.

These findings are outlined in Table 2.1 and presented in further detail below with quotes from students and images of their graph analyses.

Table 2.1: Overview of themes and subthemes in students' analyses, underlined themes and subthemes denote a written section in the results.

Themes	Subthemes	Description
Establishing Frame	Prior Knowledge and Experiences	Students draw upon prior knowledge and experiences that are activated by data within the graph and prompt. These can help activate and guide the student's frame
	Anchors	Key pieces of data and objectives of the student help students form a frame to guide their analysis
	Multiple Frames	Some students approach the task seeing multiple perspectives or interpretations of the same prompt
Observing and Interacting with Data	Relevance of Data	Students' frames help "filter" incoming information from the graph, helping them identify what graphical data is important and what is irrelevant to their analysis
	Comparisons	Students make different kinds of comparisons in graphical data to gather information for their decision-making
Data-Frame Interactions	Data-Based Evaluations	Students used objectives and anchors from their frame to evaluate their choices for the task
	Data Affecting Frame	Students consider information outside of the task prompt and embed new information into their frame, producing a small change in the frame

	Frame Changes	Students' frames changed in one of two ways: considering new information outside of the graph and considering alternative perspectives

Establishing the Frame

To begin, students called on prior knowledge and experiences that could help construct a frame for them to engage in the task. Many of these experiences involved some sort of work in the laboratory. Throughout his analysis, Evander often mentioned wanting a pure gold product to fulfill the task. When asked about this, he specifically recalled his summer work in a research lab:

"I worked in a research lab where drug purity is paramount because you don't want stuff messing with drug delivery. So I guess that was kind of in my mind." (Evander)

Even though he did not reference this research lab experience right away in his interview, it is likely that he unconsciously was drawing upon the knowledge he gained working in the environment to help form a frame to complete the task.

Not all students drew upon prior knowledge and experiences that were directly related to the task at hand. Many students brought up everyday connections to the task that helped them make sense of the task. In addition to this, some students attempted to use conceptual knowledge that was unrelated to the task. For example, Hector attempted to use his prior knowledge of kinetics to make sense of the task:

"One thing I'm like trying to incorporate would be like rates of reactions… I feel like if there were more materials other than the HCl, the gold, and the PA, then it could probably slow the reaction down." (Hector)

Here, Hector is attempting to use the concept of kinetics to help inform his analysis. In the interview, he mentioned that that it was a subject covered in his chemistry lecture, but that he had not understood it as well as other concepts. This was particularly interesting as there was no mention of reaction rates in the task's prompt, nor was there any data provided that related to reaction rates; his ideas solely came from his prior knowledge and experiences in class. Hector's focus on the molecular level of the extraction phenomenon demonstrates a relatively sophisticated level of reasoning for the task; however, none of the data presented related to rates of reaction, nor did the task ask him to consider that for his response. Although the concept of rate of reactions helped Hector initially to begin activating a frame and guide him to an answer, he later changed his response during the peer review portion of the interview. This was likely because the prior knowledge Hector had activated was limited in what it could help him make sense of.

Accessing prior knowledge and experiences for the task helped students to activate anchors for their frame and guide the frame overall. The anchors helped to define the goals or objectives that students set out to accomplish during their task, as shown by Gregor:

"Oh, I saw some keywords in the, in the question…you know, 'maximal amount of Au with minimal amounts of waste.'…So when I see things like that, I'm thinking that the problem that we're trying to solve here is one of there's a specific name for it, but it's, it's getting the most of what you want and the minimum of what you don't. Optimization, I believe." (Gregor)

Here, Gregor is focusing on keywords in the prompt and connecting them to previous optimization problems he has encountered before. The combination of the "maximal" and "minimal" terms seemed to activate his previous experiences with optimization. This allowed him to begin to establish anchors that helped guide what data he needed to consider. Gregor began to hint that his one anchor for him involved searching for an area of the graph that satisfied having a maximum of gold while still maintaining a relative minimum of other tin and iron.

Overall, anchors helped to define the goals or objectives that students set out to accomplish through their analysis. They gave direction for the student's frame to follow. Because anchors were so embedded in a student's analysis, it was often difficult to directly elicit from a student's interview. However, some students verbalized this aspect of their frame, as Ben did:

"So the first [step] is find the most Au that I can get. The second one is find the least Sn and Fe that I can get. And then the third one would be like now find the point where you can achieve both of the two steps the best way you can." (Ben)

However, not every student formed a guiding frame immediately when they began the task. Some students began by acknowledging that different concentrations of acid would fulfill different interpretations of the task, which seemed to show they were approaching the task from multiple perspectives. It is possible that these students were tracking multiple frames, one for each perspective they considered. Take Bruce:

"The biggest one, I kept going back and forth between either two or four molarity, just because I don't know how important it is to get rid of waste. Like if the main goal is to extract gold or the main goal is to minimize waste, I think that could depend on, whatever the goal is would change if you want to use two or four molar hydrochloric acid." (Bruce)

Bruce recognized that there could be multiple interpretations for the task, with each having its own set of objectives to fulfill. The "main goal" for one frame was to extract as much gold as possible, making 4 M an ideal choice. While the other frame he tracked was to avoid extracting other metals with the gold, making 2 M an ideal choice. Although Bruce saw the merit in both interpretations of the task, he ultimately chose one to pursue for the remainder of his analysis.

Observing and Interacting with Data

Once students had a frame, they began to use it to navigate the different information in the graph. The frame played a key role during this part of their data analysis, as it allowed them to filter incoming information from the graph, such as Gregor explains when describing his analysis:

"I'm looking around and I'm trying to figure out, okay, so what of this information is relevant? And that's why I went and I found our items we're actually looking for and try to ignore the rest of this to some degree." (Gregor)

Here Gregor points out that there is key information that is relevant for his analysis. His frame allowed him to solely focus on the gold, iron, and tin curves during his analysis as those were the metals of interest for his interpretation of the task. The rest of the data displayed on the graph was filtered out and ignored, as they were not part of his frame when engaging in the task. Gregor's frame also cued him to specific spaces of the graph:

"It simplifies the problem solving if I realized that all I really have to do is figure out where this is maximum ***circles maximum point on gold curve*** *and then look and see what's going on with my other sort of secondary criteria. And then if I look at that and I go, okay, well they're pretty high here* ***circles tin and iron points at 4 M***. *Then I go, well, where are they low? And I see, well, how much did we really relatively change?"* (Gregor)

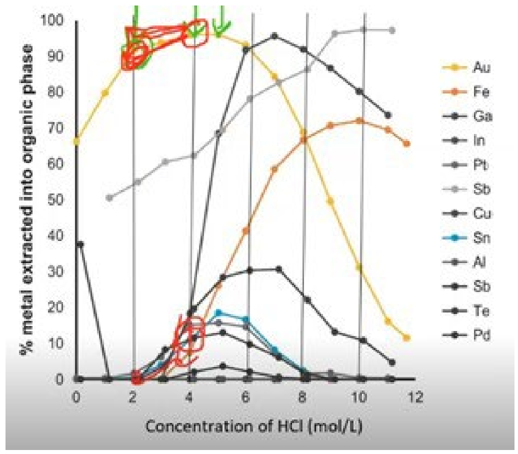

Figure 2.3: Screen capture of Gregor's graph

Here, Gregor is cued by his frame to narrow down his search of the graph to the space in which gold has reached a maximum. Once this has been located, he mentions his second objective to fulfill, which for Gregor's frame meant finding a spot with relatively low extractions of tin and iron. He then navigated to the bottom space of the graph to search for the tin and iron lines where he knew the extraction would be lower.

Following the identification of relevant data and important areas of the graph, students began to make various types of comparisons. Many of these comparisons were between a specific metal or set of metals at two different points in the graph, such as Jemma demonstrates when comparing the gold extraction at 0 M and 2 M: *"The amount of gold extracted [at 2 M] is like 25% higher than at zero where it's only 65."*

Another type of comparison made by some students was gauging the relative differences of all the metals' extractions at a given concentration of HCl. Some students did this by creating a "ratio" of the extraction of gold to the extraction of tin and iron, such as Colette did:

Interviewer: *"I wanna go back to a point you mentioned earlier when you were analyzing the graph, you talked about this idea of ratios. Can you explain that a little bit more for me?"*

Colette: *"I guess the ratio you want as much gold as you can get with having as little of the iron and tin as you can get, so like the farther apart they are. To me, if you have a 90:10 ratio, that's better than a 70:30 ratio. If you're only getting 70% of your gold, but you're getting 30% of the other metals."*

The ratio of different data points that Colette had constructed helped Collete to consider more information simultaneously. Students who constructed ratios for comparisons condensed graphical data from both the gold and the iron and tin data into a single piece of information for a given concentration of HCl. By constructing a ratio for each possibility of HCl, students could then compare each ratio and evaluate how well each choice had fit their frame's objectives via a data-frame interaction.

Data-Frame Interactions

There were two main ways in which graphical data and students' frames interacted within this task: data-based evaluations against a frame and data affecting a frame.

In a data-based evaluation, students weighed the different options of HCl outlined in the task by first obtaining some sort of relevant data from the graph. This data could be singular in nature wherein students use one point on the graph, such as a peak, or it could be more complex and involve comparisons of points or multiple points at once. Students then used this data to help assess how well the different HCl concentrations fit the objectives of their frame. If a concentration of HCl fulfilled their objective(s), they qualified it as an option for their final choice. If a concentration of HCl did not fulfill an objective or violated some aspect of their frame, the concentration was then disqualified and no longer considered an option for the final choice. To illustrate this, consider an excerpt from Gregor's analysis:

"The first thing that comes up is 4 M. 4 M comes up because that's where the gold maximum is…And then when I think about it and I decided that maybe we don't want all of this waste right here, I start looking for the minimum of the tin and iron and that's at 2 [M]." (Gregor)

Gregor begins by first noting that the peak of gold is at 4 M. The peak signifies the largest extraction of gold which aligns with his objective of achieving a relative maximum amount of gold. This qualifies 4 M as an option for Gregor to further consider against the other objectives of his frame. The next objective he considers is achieving a minimum extraction of tin and iron. Still considering 4 M as an option, Gregor then checks the tin and iron points at 4 M to see if they fit his objective. After noting that the tin and iron extraction is much higher at 4 M compared to the other options, Gregor decides to disqualify 4 M as an option for the task because it failed to meet one of his frame's objectives. Gregor then continues his analysis to examine 2 M HCl as a potential option as it seems to better fit the objective of a minimum extraction of iron and tin.

The other type of data-frame interaction involved data within the graph affecting a student's frame. This was a far less common occurrence, and students' frames experienced relatively small changes from the data. The students who had their frames affected by the data in the graph seemed to consider information that was not outlined in the task prompt and embedded the new information into their frames. For example, Fernald's analysis considered a metal that was not outlined in the task:

"Well, the ideal molarity would be around the relative maximum for the gold, and it would also be around the relative minimum for all of these other metals, especially with Sb." (Fernald)

Fernald seemed to experience a very small frame change upon beginning to explore the graph. As mentioned in the methods section, the prompt only asked students to consider gold, iron, and tin. These were also the only lines in the graph that were displayed in color to help students more easily identify them. Because of this purposeful design of the prompt and graph, Fernald likely began the task only considering these metals. However, in exploring the graph during his interview warm-up questions, he noticed the line for antimony and used it for the remainder of his analysis. Fernald may have embedded antimony into his frame as it was a relatively visible line on the graph, being lighter in color and situated above the other greyed-out metals in the graph. Despite frequent reminders from the interviewer, Fernald continually referenced antimony throughout his analysis, which suggests that his frame considered this metal in addition to the others. Specifically, I infer that he incorporated achieving a relative minimum extraction of antimony into his original frame objective of achieving a relative minimum extraction of tin and iron. Even though including antimony in his frame did not cause his analysis to differ drastically from other students in the end (he chose 2 M as his final answer, as did most other students), it illustrates an instance in which a student's frame and data interact within their analysis.

<u>When Frames Change</u>

Many students changed their answers at some point during their analysis. The change in their answer likely reflected some sort of change in the students' frames. Students often seemed to have additional or entirely new objectives for their analysis to follow, which resulted in new answers for the task. I observed these changes in answers in two different scenarios, considering new information outside the graph and considering alternative perspectives.

Frame changes that resulted from considering new information outside the graph occurred when students reread the prompt at some point. These interviews began with students either not fully reading or remembering the full prompt before starting their analysis. Students began by choosing 4 M and justifying their choice by stating the gold or all metals were highest at that point, such as Kit illustrates: *"I guess in all, I would say 4 M, because it seems like based off the 4 M in general, there's more percent metal extracted than from the other ones."* Because the students justified their answers without mentioning any other aspect of the prompt, the researcher would reread the prompt to ensure the students had fully understood the directions. Immediately after rereading the prompt students switched their answer to 2 M, often saying they misunderstood the question being asked before or that they understood the task better after rereading the prompt. The rereading of the prompt helped students to form new frame objectives that helped to guide their analysis of the graph, such as demonstrated with Kit:

Interviewer: *"So what new pieces of information do you think that you used in order to reach your conclusion?"*

Kit: *"Oh, like based off of like the new, like reading the paragraph again? I think that the key things for me was like the maximal amount of Au and the minimal amount of waste. I think that originally I had thought this [maximal amount of Au], but I wasn't considering the amount of waste."*

Originally, Kit's frame only contained the objective to find as much gold as possible within the range of concentrations given. This objective guided her to seek out the highest point for the gold line, which happened to be at 4 M, qualifying 4 M as an option to fulfill the task. Upon rereading the prompt, Kit identified an additional objective for her frame to use: achieving a minimal amount of waste. Having an additional objective in her frame guided Kit to new data within the graph that she had not considered relevant before. This new information then allowed her to consider a new concentration of HCl which progressed her analysis further into a new answer for the task.

In addition to experiencing frame changes from considering new information outside of the graph, some students also experienced frame changes by considering alternative perspectives. Most students who incurred a frame change in this way did so during the simulated peer review portion of the interview. During the peer review, students compared their own work to that of a hypothetical peer. After this comparison, some students stated that they had low confidence in their answers or were feeling less confident overall. A few students even admitted that they thought their answer was wrong. To alleviate this, the interviewer gave the student an opportunity to make changes to their answer, which gave students the chance to choose a different concentration of HCl to fulfill the task. After students had changed their answers and explained what made that answer more appropriate, the interviewer asked the student what had motivated their change in answer. Some of these students explained that the peer review offered a new perspective that seemed better than their own. For example, consider Violet after reviewing the 0 M sample response:

"This one's pretty convincing, different after reading the first student and comparing it to mine. It's kind of like seeing it from someone else's perspective. [It] just kind of puts it into a better perspective. And I feel like that makes a lot more sense than my answer…It makes a lot of sense that there's actually zero waste at 0 M and you still get a pretty good amount of the gold and no waste at all." (Violet)

The peer review prompted Violet to compare against the 0 M sample response. The comparison exposed Violet to a perspective of a frame different from her own that was more appealing. This new frame had slightly different objectives from her original frame, wherein the extraction had to achieve an absolute minimum extraction of tin and iron rather than a relative minimum. This disqualified her original answer of 2 M, which could be interpreted as having a relative minimum extraction of tin and iron (tin being slightly above zero and iron at zero) but not an absolute minimum. The only HCl concentration that achieved an absolute minimum of iron and tin was 0 M, so it qualified as the best option for Violet's frame.

Although the majority of frame changes from considering alternative perspectives arose from the simulated peer review, one student, Ariel, experienced a frame change solely by thinking of another person's perspective. Ariel began her interview with 4 M HCl as her answer. When asked to explain what had made this the most appropriate answer Ariel described her definition of minimal for the task: *"So that's kind of how, I guess I would think of it where you can't count something as minimal until it is actually present."* Using this definition, the only concentration that was qualified to fit her frame was 4 M, as both iron and tin had extractions above zero.

Intrigued by this answer, the interviewer asked Ariel to consider how the scientists behind the experiment would approach the same task. Ariel began to express ideas surrounding reducing the error of the experiment, and when asked to specifically describe what the best choice would be for the scientists, she said, *"I would say when you get the most amount of gold extracted, and then when you probably just have as little possible of tin and iron."* Here, Ariel is beginning to recognize a frame different from her own, the main difference being each frame's objective involving the iron and tin extraction. Ariel's frame defined a minimal extraction as an extraction in which both metals had a presence, whereas the new "scientist" frame defined a minimal extraction as having relatively little to no iron or tin present. Following this moment in the interview, the interviewer asked Ariel what perspective she wanted to take for the task. She admitted to feeling indecisive at this point but decided to choose 2 M HCl as her answer. After going through her analysis with the interviewer, Ariel brought up an experience from her general chemistry laboratory course that reminded her of the interview task, describing it as such:

"Well, we do like one where we mix like caffeine with HCl and like water or something…I think it's kind of similar to this one where it has like different layers…So I was just kind of thinking about that, like visually when I was doing this, cause it was kind of similar." (Ariel)

Ariel's frame change was unique in that she did not need to read another written analysis to consider this alternative perspective. Upon being asked to consider a more science-based perspective, she seemed to activate a new frame that could be used to complete the same set of directions in the task. This new perspective also seemed to help Ariel establish connections between her previous experiences in an experiment she had performed in the laboratory to the metal extraction task at hand. This in turn helped her to visualize and better make sense of the phenomenon, leading to more productive sensemaking for the task.

Discussion and Implications for Research and Practice

This study outlined the ways in which a student's prior knowledge and experiences interact with their data analysis. I specifically demonstrated that a student's prior knowledge and experience played a part in activating and developing a frame that could be used to analyze the graph. Upon approaching the task, students could draw upon relevant knowledge and experiences that helped guide them through the task. For example, one student, Evander, specifically brought up his experience working in a drug research lab as something that he used to help him navigate the task. Although his lab work did not directly mirror the task, it was similar enough for him to use as a reference. Evander used the idea of purity from his research to help identify priorities for his frame objectives to complete and support his reasoning. Slominski and colleagues interviewed several biology, physics, and engineering faculty with a fluid dynamics task in two different contexts: blood vessels and water pipes (2020). Some biology faculty used vocabulary and conceptual knowledge reminiscent of cardiovascular systems when answering questions for the water pipe task. Even though water pipes do not directly relate to cardiovascular systems, they do share the same underlying concept of fluid dynamics. Like Evander, the biology faculty used their prior experiences and prior knowledge as a reference and frame to help navigate new problems.

Not all prior knowledge and experiences were helpful for students. Consider Hector's interview; even though he was taking a rather sophisticated approach thinking of how the rate of the extraction could be affected by what metals were being extracted at a given concentration of HCl, he did end up changing answers and drawing upon different sources of knowledge to inform a changed frame later in the interview. Something in the interview questions or task prompt likely activated his conceptual knowledge of kinetics from class, and he decided to use it to inform his decision-making. Hector then seemed to struggle to connect the concept any further to the task, so he further sought other knowledge and experiences to inform his frame. This shows an example of a less productive piece of prior knowledge. Hector's knowledge was conceptually sound, but it had limited use for the context of the task. Hector saw the lack of utility in the conceptual knowledge, so he drew upon other ideas to inform his reasoning. Experts do this as well in their reasoning. Slominski and colleagues highlight one biological expert's attempt at using relatively sophisticated biological conceptual knowledge to explain a phenomenon (2020). The expert was trying to use capillary physics as a conceptual resource but was unable to connect it to the task. Recognizing this, the expert then drew upon their previous experiences with syringes and used this to inform and shape their frame.

I also observed students draw upon their everyday knowledge and experiences to help complete the task. In one example, Ariel began her interview by defining minimal as the smallest amount that was not zero. This definition informed one of the objectives for her frame, which in turn helped her choose her initial answer of 4 M. Other studies have shown this to occur during chemistry students' data analysis. Heisterkamp and Talanquer observed their participant using examples from his everyday life to explain differences in chemical phenomena (2015). Experts also use everyday knowledge and experiences to help them reason in unfamiliar contexts as well. In their dual-context fluid dynamics task, Slominski and colleagues found that some experts used a frame informed by everyday knowledge to help reason through the task situated in an unfamiliar context (2020). For example, one biologist used an outdoor water hose in their reasoning for the water pipe context. Their experience with putting a finger over the end of a water hose gave them a reference to complete the task. In all these examples, participants accessed prior knowledge from their everyday experiences that can help inform their frame for a task. There is likely some aspect of the task that activates this everyday knowledge or experience, which then serves as a frame to help the participant make sense of their task.

This work provides additional evidence that novices and experts actually undergo similar sensemaking processes (Klein et al., 2007). Both students and experts access some sort of prior knowledge or experience to help establish a frame for the sensemaking process. The key element that seems to vary in sophistication is access to relevant prior knowledge and experiences, where experts often have richer mental models and a larger knowledge base. Those who have access to prior knowledge and experiences that share similarities to the phenomenon or target similar concepts can engage in more sophisticated and productive reasoning.

To ensure that students are equipped to engage in sophisticated and productive data analysis in the classroom, instructors should consider how they are activating students' knowledge and experiences that could help students productively engage in data-based reasoning. Situating tasks in a variety of different contexts and experiences could serve to expand learners' knowledge base, thereby increasing the chances of activating students' prior knowledge and experiences that could productively frame their data analysis. Additionally, scaffolding to explicitly cue students' relevant prior knowledge and experiences can support productive engagement. The sensitivity to scaffolding was evidenced by Ariel who modified her response simply by being prompted to consider the perspective of the scientists conducting the experiment.

Overall, the participants were proficient in navigating the graph. With the scaffolding employed for the task, students did not seem to have any problems comprehending the graph and its features. Students could easily identify local maxima and minima, compare points and slopes, and interact with the surface features of the graph. Students' competency in comprehending these surface-level graph features aligns with much of the science education literature on graph comprehension in science courses (Ivanjek et al., 2016; Potgieter et al., 2008). The ease with which students navigated graphs and their features suggests they have rich graph schemas that help them to synthesize and comprehend information from a graph (Pinker & Feedle, 1990). These schemas can serve to automate graph reading for students, which helps to explain why there was very little variation in students' graph comprehension and primarily variation in what information students used from the graph.

In addition to considering how one's prior knowledge and experiences interact with graphs, there were also different ways in which a frame changed during our participants' data analysis. In all three ways, students noticed something that then altered their frame in some way. There was a variation in the degree of these frame changes, depending on how the new information aligned with a student's current frame. New information that produced smaller frame changes generally fit with the student's original frame. Smaller changes in analysis occurred when the new information did not alter any objectives within a student's frame. In Fernald's analysis, there was a very small frame change early on in his analysis. Initially, his frame had him consider only gold, tin, and iron data in the graph, as they were the only metals outlined in the task. After viewing the graph though, Fernald incorporated antimony into his frame by expanding upon his minimal extraction objective. This is an example of elaborating one's frame in Data-Frame Theory, in which the core tenets of the frame remain integral, but the frame is changed to accommodate new data that was not explained or considered beforehand (Klein et al., 2007).

To produce more significant changes in a student's frame, the student's frame needed to be challenged in some way. One way in which this occurred was by engaging students in a simulated peer review to have them consider an alternative perspective. In previous work, I used social comparison theory to model how students engaged in the simulated peer review (Berg & Moon, 2022). Here, it is likely that the simulated peer review offered students an opportunity to compare their frame to that of the sample response they were reviewing. From this comparison, students evaluated how well their frame fulfilled the task. If students were not satisfied with how their frame completed the task, they could adapt it or use a different frame entirely.

Throughout the simulated peer reviews, students demonstrated proficiency in comparing their frames to that of the sample responses. This was especially prominent for students who changed their frame after reviewing the sample response. To change their frame in this manner, students decentered from their original frame to consider another. Decentering is defined as the process of recognizing and understanding different perspectives from one's own (Piaget, 1955). Some students directly verbalized that reading the sample responses offered new perspectives they had not considered, such as Violet did when reviewing the 0 M sample response: *"It's kind of like seeing it from someone else's perspective. [It] just kind of puts it into a better perspective."* Here, Violet recognizes that the 0 M sample response has a perspective different from her own, and then acknowledges that the sample response's frame fulfills the task better than her own. By recognizing the alternative perspective and then adopting it as her own, Violet shows evidence of decentering from her original frame.

The findings provide further evidence that not only does empirical data shape a student's sensemaking during data analysis and interpretation, but also other sources of data such as socially obtained information. Previous work in physics education has found that frames can be influenced by social cues (Gouvea et al., 2019). In the context of this study, social cues and social information were accessed by engaging in the simulated peer review. Through the peer review, students were exposed to alternative perspectives that they could consider and potentially adopt as their own for the task. Further work could and should consider how a student's frame influences how they may engage in a peer-based classroom activity. This work is based in peer review, but there likely are similar occurrences in other peer-based classroom activities such as peer-led team learning. This is an opportunity for further research to a) uncover how peers prompt modifications to a learner's frame and b) elucidate how scaffolding and intangible cues may serve to activate specific frame components, or even prompt reflection on and help aid in productive modification of frames.

I encourage classroom practitioners to consider how they are attempting to elicit sensemaking from students in tasks. Instructors should account for how the information in the task, such as empirical data and prompt instructions, may activate specific frames with which students make sense of the data. When designing the task, instructors can also think of how to leverage certain prior knowledge and prior experiences that relate to the task. These both can help guide students in forming an appropriate frame to make sense of the task productively. Finally, instructors can consider how other sources of information, such as alternative perspectives of peers, may influence students' frames in their sensemaking process and overall engagement in a task. Other sources of information, like reviewing peer's work, could have the power to challenge students' less productive or relevant ideas.

Conclusions

The present study shows three themes and one sub-theme related to how undergraduate chemistry students used their prior knowledge and experiences when analyzing a line graph for a task. The themes are shown in Table 2.2.

Students first established a frame to help make sense of the data. To do this, students activated relevant prior knowledge and experiences that helped to establish goals for their analysis to accomplish. Students may even have started with multiple frames before deciding which frame best suited the task.

After establishing their frame, students observed and interacted with the data in the graph. The frames helped students to identify important data that would help them to accomplish their goals. Students also compared different data features or sets of data features to gather information for their decision-making for the task.

Once students had made observations of the data, students had different data-frame interactions. Often, students used graphical data to evaluate one of the three options for the task. The students' evaluations helped students to decide which task option best fit the goals of their frames. A few students experienced another type of data-frame interaction in which the students' observations of the data incorporated new additional goals into their frames.

The final theme explored is when students' frames experienced changes. This was another form of data-frame interaction; however, these changes resulted from students considering information outside of the graph. These changes occurred after students re-read the task prompt and considered information they had overlooked when they previously read the prompt. Alternatively, some students made changes to their frames when considering alternative viewpoints they were exposed to through a simulated peer review.

Table 2.2: Description of themes and subtheme of students' analyses

Theme or Subtheme Explored	**Description**
Establishing Frame	Students activate prior knowledge and experiences that help to establish goals for their data analysis to accomplish. Students may begin with multiple frames but eventually decide on one to use for the task.
Observe and Interact with Data	Students' frames help them to identify what data features must be considered in their analysis. Students compare data features to garner more information for their decision-making.
Data-Frame Interactions	Students use graphical data to evaluate which of the three options for the task best meets the goals of students' frames. The graphical data may also cause students to incorporate additional goals into their frames.
Frame Changes Resulting from Data-Frame Interactions	Some students' frames changed when re-reading the task prompt or when being exposed to alternative viewpoints through the simulated portion of the interview.

Limitations

This study's characterizations of students' data analyses are highly contextualized in the design of the task and research methods used.

First, the task utilized a familiar data representation for students. It is extremely likely that undergraduate students have encountered graphs and used them at some point before engaging in this task. Familiarity with analyzing and using graphs affords robust graph schemas for students to use, which made the navigation of the graph relatively effortless for students. Throughout the interviews, students had no difficulties navigating the graph to find certain information, make comparisons, or in comprehending the graph in nearly any manner. Other kinds of data representations, both domain-specific representations (such as NMR spectra) and domain-general representations (such as data tables) are very likely to produce different sensemaking from students.

It is also important to acknowledge that the design only required students to consider one set of data. For this work, I simplified the task so that students only needed to consider a graph of extraction values, which narrowed students' focus to three specific metals. The task was designed to be simple so that the analysis could both focus on characterizing students' sensemaking process and identifying students' prior knowledge and previous experiences used in their sensemaking. Tasks that use multiple sets of data will produce more complex sensemaking and potentially more complex frames. This would be especially likely for tasks in which the data sets contradict one another in some way.

The content knowledge required to navigate the task also poses a limitation. The task was purposefully scaffolded to be accessible to chemistry students enrolled in first- and second-semester general chemistry courses. Much of the underlying chemistry content was removed so that students could engage in the task with relative ease. Tasks that would require more content knowledge to engage in will produce much more complex sensemaking from participants. It is also likely that requiring more content knowledge would produce more variation in sensemaking from participants as well.

Finally, the use of interviews does pose a limitation in generalizability to broader contexts; however, the choice to use interviews enabled deep examination of students' thought processes during their analyses as well as the identification of prior knowledge and experiences that influenced their analyses. During data collection, saturation was reached at eighteen interviews as students in the later interviews used no new reasoning or perspectives that had not been used before (Nelson, 2017).

CHAPTER 3

PROMPTING HYPOTHETICAL SOCIAL COMPARISONS TO SUPPORT CHEMISTRY STUDENTS' DATA ANALYSIS AND INTERPRETATION

Abstract

To develop competency in science practices, such as data analysis and interpretation, chemistry learners must develop an understanding of what makes an analysis and interpretation "good" (i.e., the criteria for success). One way that individuals extract the criteria for success in a novel situation is through making social comparisons, which is often facilitated in education as peer review. In this study, I explore using a simulated peer review as a method to help students generate internal feedback, self-evaluate, and revise their data analyses and interpretations. In interviews, I tasked students with interpreting graphical data to determine optimal conditions for an experiment. Students then engaged in social comparisons with three sample responses that I constructed and compared these samples to their own. I present a model informed by social comparison theory that outlines the different processes students went through to generate internal feedback for their own analysis and response. I then discuss the different ways students use this internal feedback to determine if and how to improve their responses. This study uncovers the underlying mechanism of self-evaluation in peer review and describes the processes that led students to revise their work and develop their analysis. This work provides insight for both practitioners and researchers to leverage student's internal feedback from comparisons to self-evaluate and revise their performances.

Introduction

Reforms in science education have called for the integration of science practices (i.e., the ways scientific knowledge is generated) into science instruction (National Research Council, 2012). Although there is consensus on the need for integrating these practices in the classroom (Cooper et al., 2015; National Research Council, 2012; Singer et al., 2012), the science education community continues to investigate methods to support students' competency in science practices. Recent research has found critique to be an essential component of competency in science practices (Ford, 2012; González-Howard & McNeill, 2020; Osborne et al., 2016). Of the eight science practices, research on student engagement in data analysis and interpretation has uncovered a multitude of challenges many students face.

Of the documented challenges, many students begin to experience difficulties when working with empirical data. Students may fail to differentiate important data from unimportant data (Jeong et al., 2007). Students may also focus on the surface features of data and ignore salient features that target the given phenomenon (Heisterkamp & Talanquer, 2015; Kanari & Millar, 2004). This can lead to students uncovering less relevant patterns in the data that do not effectively target the phenomenon (Zagallo et al., 2016). Focusing on these surface-level patterns in a dataset may lead to students missing the relevant scientific concepts. Students also face challenges when connecting patterns back to the target phenomenon. Many students will form conclusions with misconstrued reasoning or neglect to use scientific reasoning entirely when connecting uncovered patterns from datasets to the target phenomenon (Becker et al., 2017; Heisterkamp & Talanquer, 2015).

To overcome these challenges and support students in developing competency in data analysis and interpretation, I propose peer review as a method to help students develop evaluative judgment in their data analysis and interpretation. In this study, I simulate peer review to explore how critiquing peers' work helps learners develop evaluative judgment. Understanding how students evaluate their own work when giving feedback to others can inform and improve peer review practices in the classroom. Additionally, it can offer a practical approach to supporting undergraduate students' development of competency in science practices.

Background

Data Analysis and Interpretation

Data analysis and interpretation is one of eight science practices outlined in the Next Generation Science Standards (National Research Council, 2012). Practicing data analysis and interpretation often involves making sense of a visual representation, such as a graph or a table, and using it to form a conclusion. There are several processes students engage in to interpret data. First, the visual representation of data must be "decoded" where the information that is embedded is extracted (Carpenter & Shah, 1998; Glazer, 2011; Shah & Hoeffner, 2002; Zagallo et al., 2016). The difficulty of this step can vary depending on the kind of representation that is being used and the amount of information that is embedded within it (Glazer, 2011; Shah & Hoeffner, 2002). From here, relevant patterns within the data must be identified (Carpenter & Shah, 1998; Glazer, 2011; Shah & Hoeffner, 2002; Zagallo et al., 2016). This step can prove difficult for many students as studies have shown that students may selectively use data or struggle to differentiate important features of data from unimportant ones (Jeong et al., 2007; Kanari & Millar, 2004). By focusing on the less important information, students may uncover irrelevant patterns within the data (Zagallo et al., 2016). Focusing on less relevant data and patterns ultimately proves problematic when students must then tie the patterns back to the target phenomenon to form a conclusion or explanation (Carpenter & Shah, 1998; Glazer, 2011; Shah & Hoeffner, 2002). When the claims and explanations constructed from less relevant data do not effectively target the phenomenon for which the data has been collected, students may risk missing the relevant scientific concept entirely (Lai et al., 2016; Zagallo et al., 2016). For example, Zagallo and colleagues found that some groups of undergraduate biology students in a transformed Cell and Developmental Biology course became distracted by less relevant data during a classroom problem set (2016). Although the students did eventually shift their focus to the relevant data, they did lose valuable class time and needed guidance from an instructor to lead them to the relevant scientific concept.

Similar challenges have also been identified for data analysis and interpretation in chemistry contexts. Like in many other domains, chemistry students will often rely on surface features or less relevant features of data representations and models to form conclusions or construct explanations (Becker et al., 2017; Heisterkamp & Talanquer, 2015). In addition to this, many chemistry students will use misconstrued reasoning or neglect to use reasoning entirely when engaging in data analysis and interpretation (Becker et al., 2017; Heisterkamp & Talanquer, 2015). In a case study investigating the major types of reasoning general chemistry students use when engaging in data analysis and interpretation, participants relied on "hybridized" reasoning and mixed intuitive knowledge with their chemical knowledge when producing explanations (Heisterkamp & Talanquer, 2015). In another study investigating how students construct mathematical models to describe rate laws from empirical data, many of the students did not connect the mathematical model they had produced to the actual trends in the data (Becker et al., 2017). Becker and colleagues also found that some of the participants engaging in the data analysis and interpretation had produced conclusions without even consulting the kinetic data given to them. This is perhaps the most problematic approach to data analysis and interpretation, as the Next Generation Science Standards states that students must "present data as evidence to support their conclusions" when engaging in data interpretation and analysis (National Research Council, 2012).

The current literature in psychology, science education, and chemistry education has described how students engage in the practice of data analysis and interpretation and documented common challenges for students; however, little work has explored how to support the development of their data analysis and interpretation skills (Bolger et al., 2021; Zagallo et al., 2016) and no work has been done in chemistry.

Peer Review

Peer review offers a unique opportunity to expose students to their peers' work. The potential benefit is especially promising for tasks that require students to generate a product, as many of the science practices do (in this case, an evidence-based decision). Reviewing peers' work can help students evaluate their own work and potentially make changes to improve it. Making improvements to their work could entail incorporating new evidence or reasoning they had encountered in their peer's work, or it could even involve producing an entirely new conclusion if their peer's work is more compelling than their own. On the other hand, if their peer's work is like their own, students may develop confidence in their conclusion. In this way, engaging students in peer review has been shown to develop evaluative judgment (Nicol et al., 2014). Evaluative judgment includes an understanding of the criteria for success and quality work within a given domain (Sadler, 2010). Therefore, developing evaluative judgment is key to learning what makes data analysis and interpretations good. For science practices, evaluative judgment is part of what has been referred to as deep understanding, or understanding of the epistemic criteria of science (Kuhn et al., 2017). Deep understanding is a key learning objective for engaging learners in science practices.

The process of receiving feedback in peer review has received much of the attention within peer review literature. Receiving feedback from multiple peers can help students evaluate their work and make changes to improve the quality of their work more so than only receiving feedback from an instructor (Cho & MacArthur, 2011); however, receiving feedback from peers does not guarantee a student will make necessary revisions to their work (Finkenstaedt-Quinn et al., 2019). Students must recognize the value of the feedback they are given and make judgments on what feedback must be incorporated, while also managing affect surrounding the feedback (Carless & Boud, 2018). This process of enabling feedback uptake takes time and labor to develop for both instructors and students.

Recent work has found that the gains from receiving feedback are less than the gains from giving feedback in peer review (Anker-Hansen & Andrée, 2015; Cho & MacArthur, 2011; Ion et al., 2019; Lundstrom & Baker, 2009; Nicol & McCallum, 2021). Giving feedback appears to engage students differently than receiving feedback from others. When giving feedback, students make comparisons with their own work (McConlogue, 2015; Nicol et al., 2014; van Popta et al., 2017). The student's own product will often serve as a reference to compare against. The comparison process allows students to engage in active reflection on the task criteria and their own work (McConlogue, 2015; Nicol et al., 2014; Nicol & McCallum, 2021). Through producing feedback for others, students can generate internal feedback to inform and revise their draft to be in better compliance with their understanding of the task criteria. Students have reported that revising their draft in this way reduces the need for receiving feedback from peers, as they had already made the changes suggested to them when they reflected on their own work (Anker-Hansen & Andrée, 2015; Nicol et al., 2014; Nicol & McCallum, 2021).

To better understand how students evaluate and revise their own work when giving feedback to others, it is necessary to first consider the process of revising a written draft. Previous studies in college writing have found that when making revisions, students engage in a four-step process (Flower et al., 1986). First, they define the task, gaining a deeper understanding of what must be done in the task. This part of the review process is further supported by students self-reporting that they can take the perspective of an assessor and better understand the given standards for the task when providing feedback (Nicol et al., 2014). Second, students detect any problems that might be present in the work. To detect a problem, students must recognize the differences between the given work and an ideal work that follows the standards defined in the first step. Students will often use their own work as a standard to compare against (Nicol et al., 2014). The differences that students find between the given works will likely be the problems they detect. Once the problems have been detected, they can be further identified in the third step: diagnosis of the problem. Flower states that the diagnosis of a problem "brings new information to the task" (Flower et al., 1986, p. 41). The problem diagnosis is not necessarily essential for the revision process; however, identifying and articulating the nature of a given problem is associated with more sophisticated revisions (Patchan & Schunn, 2015). Finally, a solution strategy is offered as the final step in the revision process. A strategy may involve getting rid of a problematic portion or revising and rewriting the given task.

The cognitive processes of making revisions overlap with many of the cognitive processes associated with providing feedback for others in peer review (Patchan & Schunn, 2015). Students must be able to detect a problematic part of a work, diagnose what makes that part problematic, and then determine a solution strategy to improve the work. In addition to these processes, current peer review literature has also outlined how peer review can act as a vehicle to generate internal feedback for students (Nicol, 2020; Nicol & McCallum, 2021). Because students use their own work as a benchmark to make comments on other's work, the resulting comparisons will promote active reflection on one's own work and help generate internal feedback about their performance. Generating this internal feedback is one way that students can make improvements to their own work (Butler & Winne, 1995; Nicol, 2020; Nicol et al., 2014; Nicol & McCallum, 2021). A key step in this comparison is the explicit differentiation between a peer's perspective and one's own, or decentering (Moon et al., 2017; Teuscher et al., 2016). Decentering itself has been shown to be productive in supporting one's own reasoning and interactions with others (Moon et al., 2017; Teuscher et al., 2016).

Social Comparison Theory

Peer review is an inherently social process in which students typically engage in comparison with other's work. These comparisons are affected by how the student perceives themself in relation to others. This will also impact the internal feedback that students generate from evaluating their own work while giving feedback in peer review. I propose using social comparison theory to investigate how chemistry students generate internal feedback and evaluate their own work when giving feedback in a peer review setting.

Social comparison theory was originally developed by social psychologist, Leon Festinger, in 1954. He theorized that when placed in ambiguous environments that produce uncertainty about how to think or behave, individuals will compare themselves with others in the same situation to reduce that uncertainty (Festinger, 1954). Later research in social psychology has found that people will often engage in social comparison in situations where there are specific criteria and standards (Alicke, 2007; Greenwood, 2017; Levine, 1983; Martin, 2000;

Miller et al., 2015; Pomery et al., 2012; Smith & Arnkelsson, 2000). These comparisons serve to gauge an individual's performance and ability relative to others.

When engaging in social comparison, an individual will compare to a "target" (Alicke, 2007; Greenwood, 2017; Martin, 2000; Miller et al., 2015; Pomery et al., 2012; Smith & Arnkelsson, 2000). The target is simply the subject(s) to whom the individual compares themself to, and these subjects can be real or imaginary as long as they exist in a similar environment or situation. The individual's perception of the target's performance will determine the kind of social comparison being made. If the target's performance is perceived as superior in some way, it is considered an upward comparison. If the target's performance is perceived as inferior in some way, the comparison is considered downward. Performances that are perceived as similar are considered lateral comparisons. The direction of the social comparison is often influenced by the motivation for the social comparisons, beyond reducing uncertainty.

Further research in social comparison theory has found that there are two primary additional motivations for engaging in social comparisons: self-improvement and self-enhancement. Self-improvement is associated with upward comparisons (Dijkstra et al., 2008). By comparing one's work to a "better" model, an individual has the chance to gain inspiration or learn how to improve their own work. On the other hand, self-enhancement is associated with downward comparisons (Dijkstra et al., 2008). Individuals will engage in a downward comparison with a target that they perceive to be worse off. This aids in helping the individual improve their perception of their own work, easing the anxiety and low self-esteem surrounding their performance or ability (Dijkstra et al., 2008).

Social psychologists have argued that the classroom creates the ideal conditions for engaging in social comparisons (Pepitone, 1972). Students are motivated to improve their learning and the act of learning new material in the classroom often generates cognitive uncertainty. Therefore, students are motivated to engage in social comparison as a method to evaluate and obtain internal feedback on their performance (Levine, 1983). Some, however, have hesitated to use social comparison in the classroom due to the negative connotations of comparing oneself to others. There are underlying assumptions that engaging in social comparison could potentially cause feelings of inferiority, competitiveness among peers, and decreased motivation for some students (Levine, 1983). To minimize this possibility, I propose adjusting the conditions of social comparison within the peer review process by lowering the stakes of the comparison and having students review anonymous, preconstructed responses (Beach & Tesser, 2000).

Using social comparison theory to investigate offers the opportunity to focus on the thoughts and processes of the individual in a peer review setting. I propose using this theory as a lens to explore the mechanisms by which a student evaluates their own work and generates internal feedback while giving feedback in a simulated peer review. This specifically guided the study to answer this central research question: How do chemistry students evaluate their own data interpretations when critiquing hypothetical peers' data interpretations?

Methods

Overview

The central phenomenon the study is to investigate is how students evaluate their own work when giving feedback in peer review, specifically focusing on the cognitive processes that students go through to make changes to their work (Creswell & Poth, 2016). Semi-structured interviews were used to allow for a systematic approach in interviews with the added flexibility of probing questions.

Interview Protocol

The interviews consisted of two stages. The first stage of the interview engaged the students in data analysis using a modified line graph shown in Figure 3.1. The data was taken from a report of an experiment performed to extract gold from waste electronic and electric equipment (Doidge et al., 2016). Students were tasked with finding an optimal concentration of hydrochloric acid to obtain a maximal extraction of gold with a minimal extraction of waste metals tin and iron. Students were given relevant experimental details to aid in their analysis and interpretation of the graph to help them choose between three different concentrations of hydrochloric acid (0M, 2M, and 4M). At the end of this stage, students produced a written response to convince someone else and explain why their concentration choice was the best.

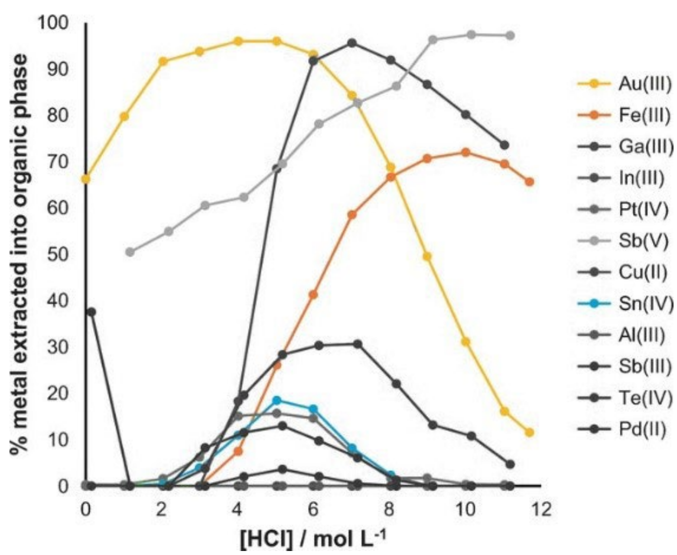

Figure 3.1: Graph modified from Doidge et al. (2016)

In the second stage of the interview, participants evaluated three sample responses. They were told that these sample responses had been generated by students participating in the same study (i.e., interpreting the same data). The three responses corresponded to the three concentrations of hydrochloric acid considered in the first stage of the task. Importantly, I constructed each sample response to include potential epistemic errors that could be made in this context (e.g., only considering one variable). Each sample response contained accurate information from the graph but used different reasoning to support one of the three hydrochloric acid concentrations. Students were presented with one sample response at a time to review. Students often began by identifying points of strength and weakness for the sample. If they did not explicitly bring up their own response at this point in the interview, they were directly asked to compare it to the sample response. This point served as the social comparison of the interview. Students generally brought up differences between the content and the quality of the responses and their own. Follow-up questions were asked as needed to elicit comparisons of both the content and quality types. After the comparison, students shared their feelings about their own responses and analysis. This point served to gauge the student's confidence from engaging in the social comparison and providing feedback to the sample. If the student stated they felt less confident or had low confidence, the interviewer asked the student what kinds of changes they would make to their own response to improve their confidence. Students also shared why they felt their confidence was affected by reading the sample. See Appendix A for the full interview protocol with the three sample responses.

Sample Selection

The study took place at a large, Midwestern university in the fall of 2020. Approval was obtained for the study from the Institutional Review Board before recruiting for interviews and students consented immediately prior to participating in the interviews. Participating students (N=18) were recruited for semi-structured interviews from both first-semester and second-semester general chemistry courses near the end of the semester. All interviews took place remotely over Zoom a week after the semester had ended.

Data Collection

During the interviews, students used the chat feature to write their own responses for the data analysis task, and later to review their responses and read sample responses. I presented the student with one sample response at a time. Interviews lasted between 45 and 80 minutes. The resulting audio recordings were transcribed via Temi.com or Zoom, and the video recordings were kept for reference in the case that students made comments referring to visual features in the graph. All data collected from the interviews were deidentified and pseudonyms were assigned to each interview participant. Participants who completed the interview were compensated with $20 gift cards for their help in the study.

Data Analysis

For this study, I analyzed interview transcripts from the second stage of the interview. In each interview, a student responded to three sample responses; thus, a total of fifty-four students' social comparisons to sample responses were collected. Two of the social comparisons were excluded from the analysis and results because students did not show evidence of engaging in the social comparison.

To begin the analysis, I used a combination of process coding and open coding to find patterns in students' responses (Miles et al., 2014). Process coding is a form of open coding that uses gerunds to describe observable and conceptual actions performed by the participants in the study (Miles et al., 2014). All process codes and other open codes produced resulted from students' own words describing their actions and confidence throughout the task. Additionally, there were some codes developed a priori to describe gaps students identified within each written sample. These codes were weaknesses I had purposefully constructed into each response, and I anticipated students would identify them at some point in their interviews.

The process codes that I developed were used to describe students' actions throughout the interview (Miles et al., 2014). I began by reading through each interview to identify how students responded to each sample response. As certain actions were repeated within interviews and across interviews, codes were generated to describe the specific action. These codes related to both how students reacted toward the written samples and their own responses. Some examples of process codes from this point in the analysis include "offering constructive criticism", "dismissing sample", and "changing claim."

To investigate students' confidence, each interview was read through to see how students gauged their confidence when responding to different written samples. I coded the points when students stated an overall level of confidence or change in confidence, specifically noting if the students had stated they had higher or lower confidence. In addition to coding students' confidence, many students with lower confidence made statements such as "I don't know" or "I don't know about…" while engaging in the social comparison with the written sample. I considered these to be instances of students expressing cognitive uncertainty surrounding some element of the task. Mitigating uncertainty is one of the motivations people may have to engage in social comparison (Festinger, 1954; Greenwood, 2017; Martin, 2000; Miller et al., 2015; Pomery et al., 2012; Smith & Arnkelsson, 2000); therefore, by accounting for students' expressions of uncertainty and documenting the specific elements students expressed uncertainty about, students' thought process could be better followed throughout the social comparison.

In the next iteration of the analysis, I used axial coding to see how the different codes generated from open coding related to each other. This mainly involved relating the different process codes together to describe the general actions that students engaged in when giving feedback to the samples. I first used constant comparative analysis to sort students' responses to each sample response based on whether they found gaps in their own responses or not. The gaps were indicative of critical internal feedback the student had generated regarding their own work. From there, students were further sorted based on any changes in confidence they expressed after engaging in the social comparison with a sample response. This sorting included accounting for increases, decreases, or no notable changes in confidence. Finally, I further sorted students based on how they responded to their change in confidence. This first consisted of sorting students based on whether they made changes to their responses. Students who did make changes were then further sorted according to how they modified or planned to modify their responses.

In the final stage of analysis, selective coding was done to piece together the general actions from the axial coding to outline the processes involved in giving feedback in peer review. Actions were put in sequential order to develop a model of obtaining internal feedback from peer review with the four categories from the axial coding stage as potential paths that could be taken. Student confidence and uncertainty of their own response were also incorporated into the model as observable events to track which path a student might end up taking when engaging in peer review during the interviews.

Trustworthiness of Analysis

Through these iterative cycles of coding, a coding scheme was developed to characterize student internal feedback from engaging in peer review and the changes students made to their original responses. A researcher from outside of the project was trained on the coding scheme and then independently coded 10% of student's peer reviews in the interviews using the coding scheme. The coded subset of data was then compared and discussed between the outside researcher and me until consensus was reached. The main points of discussion concerned internal feedback for students who did not make any changes to their responses and stated they did not have any change in confidence for a given social comparison. Originally, this response category only included students who identified gaps in the sample responses but did not state a notable change in confidence or make changes to their responses. The outside researcher and I discussed how to categorize a small sample of students who did appear to actively engage in comparison with the sample responses. These students identified gaps in the sample responses but also recognized the alternative responses as valid ways to approach the task. These students did not make any changes to their responses, nor did they state that they felt differently about their responses from the social comparison. Because of the lack of change in their response and lack of change in confidence, I inferred that they did not gain any observable internal feedback from the comparison. To capture this type of reaction, the "No Internal Feedback" category was expanded to include this type of reaction more explicitly. Discussions such as these refined the coding scheme and working model. I coded the remaining data, but the other trained researcher was consulted on two interviews for an additional perspective on coding.

Results

Overview of Results

From the semi-structured interviews, I constructed a model that describes how students generated internal feedback by giving feedback in a one-sided peer review setting and how they then used the internal feedback to evaluate their work (Figure 3.2). All students began with the same process of forming their own response and continued comparing and evaluating it against a written sample, but from there diverged into different paths depending on the kind of internal feedback they had developed from the social comparison. These paths further diverged based on how students used and responded to the internal feedback they had generated from the social comparison.

In addition to outlining the processes associated with generating and using internal feedback from peer review, the model also considers how a student's confidence and uncertainty change and influence how they use any internal feedback. After engaging in a social comparison, students' confidence often changed, which seemed to relate to the internal feedback they had generated from the comparison. I observed students with lower confidence and more uncertainty in their original responses re-evaluate their original analysis. When the social comparison might have caused some uncertainty surrounding the quality of their work, many students were motivated to address their uncertainty by making changes to their answers.

I observed four different types of responses to the social comparison illustrated in Figure 3.2. Each response category is distinct based on what kind of internal feedback students generated from the comparison, their resulting confidence after the comparison, and how students responded to their internal feedback. The response categories were also tied back to the different motivations that have been identified in social comparison theory.

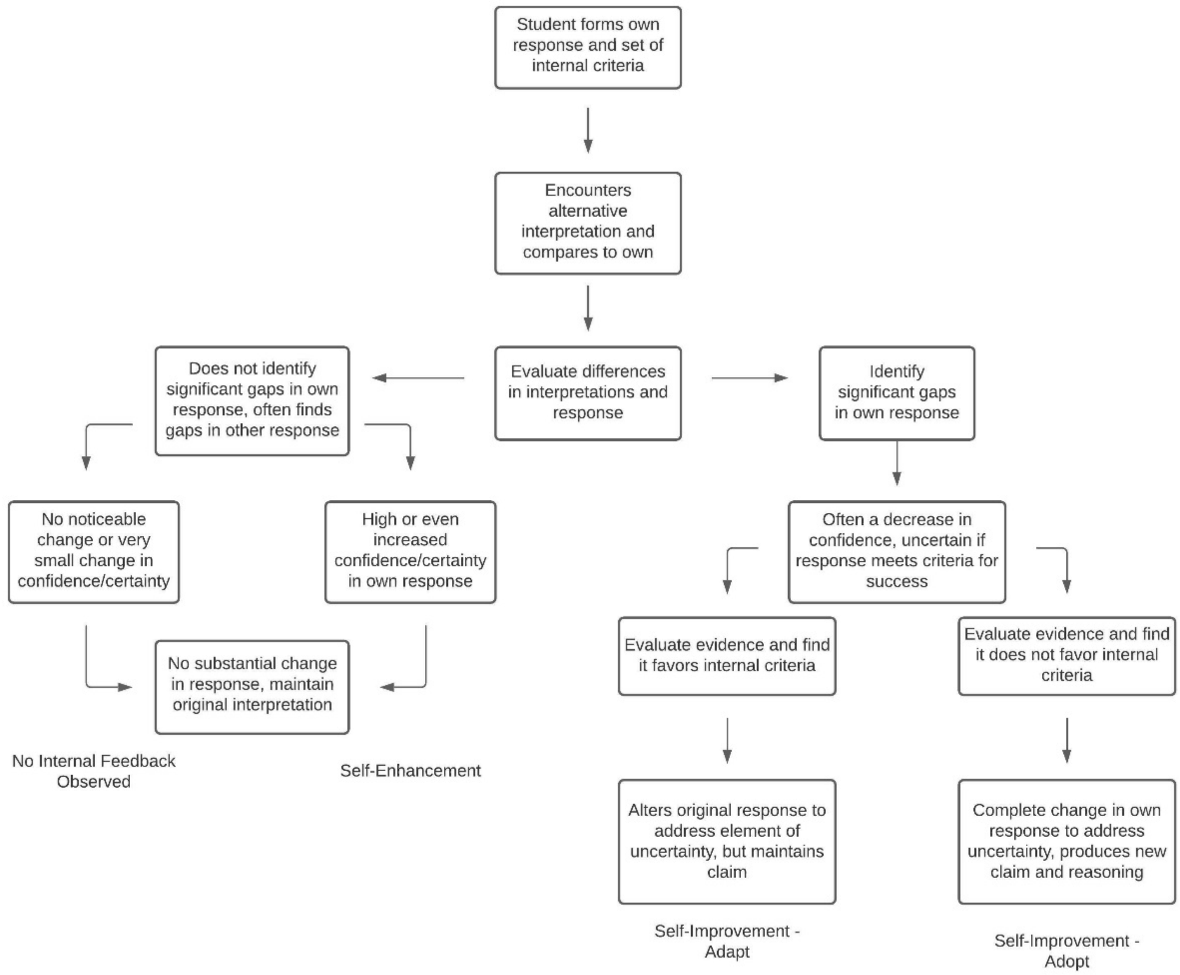

Figure 3.1: Model of different paths through social comparison, internal feedback generation, self-evaluation, and revision

Form Internal Criteria

Students first formed a response to fulfill the task. To do this, students needed to interpret the prompt from the first phase of the interview and define the criteria needed to fulfill the task. For this task specifically, students needed to form criteria surrounding what minimal and maximal meant within the context of an extraction, and then translate the interpretation to the graph to find an appropriate answer for the task. These terms were ambiguous enough that there was no universal definition for students to use, so students were required to ascribe some sort of meaning to them. In constructing meaning for his criteria, Bruce (2M), like many of the other participants, defined minimal as closely related to minimum, but not the same:

"Minimal doesn't mean the same thing as minimum, if I'm not mistaken… I would make the assumption that 1% is a minimal amount of waste, but it's not the minimum amount of waste. So 1% is a really small amount of waste, but it's not the smallest amount of waste." (Bruce)

Here, Bruce explains part of the criteria for his own response, noting the differences between a criterion of minimum waste and minimal waste within the context of the task. His answer was chosen and constructed to reflect his definition of minimal as a component of his criteria. Because of how implicit it was within a student's analysis, the criteria itself often did not surface until students engaged in the social comparison. Bruce, like many of the other students, did not fully explain what 'minimal' meant until he encountered another interpretation of the same prompt. It was through encountering an alternative interpretation of a sample and comparing it to their own that most students mentioned the standards used for their own responses.

Comparison of Sample Response

After constructing their own response according to their criteria, students then encountered an alternative response and compared it to their own. Students often made comparisons of their interpretation of the prompt to the sample's interpretation, using their own interpretation as a benchmark. For example, consider Evander's comparison of his own interpretation (2 M) to the 0 M sample:

"They [0 M response] considered the impurity as the end all be all, however much gold we extract in the end, it is what it is. I kinda met or I started with at four and then worked my way

down to two. I had the maximize gold approach and then the minimizing the impurity was kind of second hand to that." (Evander)

Here, Evander recognizes that the 0 M response had interpreted the prompt differently than he had and is able to identify how the response differs from his own. He then provides his own approach to fulfilling the task, demonstrating that his own response acted as a benchmark for the comparison. Importantly, Evander very specifically uncovers the difference between the criteria being used in the sample and his own. Evander argues that the sample author considered only one criterion: eliminating impurity; whereas Evander prioritized maximizing gold followed by considering the impurity. Evander's quote illustrates the decentering that served as the first step in comparing a sample response to one's own. While all students used their own responses as a benchmark for a comparison to the sample, some students also made additional comparisons to previous samples they had encountered in the interview. These comparisons were similar in nature to ones in which students used their own response as a benchmark, they just included more targets to compare to and these occurred later in the interviews after students had encountered multiple sample responses.

<u>Evaluation of Sample Response</u>

Students' comparisons with the sample response served as a basis for the evaluation step of the model. During this step, students assessed the sample to see how well it fit with their internalized criteria formed during the first step. When assessing the sample, students would identify different strengths and weaknesses of the sample response. Once these were identified, students would go on to determine how well the response aligned with the internalized criteria they formed from the prompt. For instance, Evander's (2 M) evaluation of the 4 M response is heavily informed by his motivation for a pure gold extraction.

"They [4 M response] focused purely on the amount of [gold] extracted and they didn't take into account the potential for impurities as the concentration [of HCl] increased. So I guess starting from zero and going to four, like when they talked about that 65 to 95, they didn't, I guess not understand, but they didn't take into account the other two compounds that are classified as waste within the question." (Evander)

Evander began his evaluation by identifying a gap in the 4 M response: the response only included information on gold. He recognized that the prompt considered two of the metals included in the task were waste and could be extracted with gold, causing an impure extraction to take place. Having a pure extraction was a criterion that informed Evander's own response for the task, so encountering a sample response that was not aligned with this criterion ended up resulting in a negative evaluation of the sample response.

Self-Enhancement

How well the sample response fit with a student's internal criteria influenced what kind of internal feedback was obtained from the social comparison. Students who did not find that the sample response fit their internal criteria typically found gaps in the response that made it weaker. This often gave students favorable internal feedback from the social comparison, as their own responses did not have these gaps making them relatively stronger; however, some students were able to obtain favorable internal feedback from social comparisons with sample responses that were similar in strength to their own, as is the case with Ben (2 M).

Interviewer: *"Okay. And how does this response [2 M response] compare to yours?"*

Ben: *"I think it's kind of on the same level. I think we're saying the same thing. I don't really see it as false. We both do the same kind of analysis and like we compare both of them while acknowledging the maximum and the minimum amounts."*

Interviewer: *"How are you feeling about your response after reading this one?"*

Ben: *"I'm feeling good because I see that someone did the same thing I did. They analyzed it the same way without any -- like it doesn't differ from mine. If this differed from mine and the conclusion was different, that would make me less confident because I can see I had an error in mine, which makes mine not correct."*

Although Ben had identified some argumentative gaps in the 2 M sample response earlier in the interview and had suggested that the response include more evidence to support its conclusion, he still viewed it as similar in quality to his own. He found that his own response and the sample had similar analyses and interpretations of the prompt that in turn validated the internal criteria for his own response. Seeing that his own internal criteria and analyses were mirrored in the 2 M sample response gave Ben positive internal feedback. Experiencing validation and higher confidence from positive internal feedback like this was indicative of a student experiencing self-enhancement from the social comparison. Students who experienced self-enhancement from the social comparison did not make changes to their response in any way; therefore, they were not considered to have been motivated to change their response. The validation they gained from the social comparison helped them to feel confident enough in the strength of their response that they likely did not feel an incentive to revise it.

<u>Self-Improvement Route</u>

In contrast, students who experienced critical internal feedback often lost confidence in their own responses, as observed in Ben's explanation. Had the sample response had a different conclusion and analysis (and presumably fit his internal criteria) Ben would have lost confidence in his own response. This is what occurred with other students in their social comparisons, as Violet (2 M) noted after her comparison with the 0 M sample response:

"*My answer [2 M] made sense to me when it was just me thinking it through. And then getting the perspective of these other two students and what they think—it just makes more sense to have absolutely zero waste and have 65% of the gold. Versus my answer, you're having 90% of the gold but you have a little bit of waste... And in the paragraph, they want to use the maximum amount of gold with minimum amounts of waste. So, it just makes more sense to have absolutely zero waste and then you know that it is just the 65% of the gold going through*." (Violet)

Upon making the social comparison with the 0 M sample response, Violet generated critical internal feedback for her own response. Even though her original response seemed to fit her original internal criteria at the time, it did not seem to align with her new internal criteria as much as the 0 M sample response did after the social comparison. Violet's internal criteria seemed to shift after being exposed to the perspective of the 0 M response. She then identified that the amount of waste at 2 M in her original response did not satisfy the "minimum amounts of waste" criterion as well as the 0 M sample response did after the comparison.

After engaging in the social comparison, students who generated critical internal feedback generally expressed doubt about the quality of their original responses. They then were given an opportunity to make changes to their response to address any of the gaps they identified in their own response. By addressing their uncertainty in the quality of their response, students demonstrated that they were motivated by self-improvement. To begin improving their response, students first evaluated the alignment between their internal criteria and their response. The results of this evaluation then went on to affect what kinds of changes students made to their responses.

Self-Improvement: Adaption

Students who found their original response still mostly fit their internal criteria maintained the essence of their response but made smaller changes. This kind of response towards critical internal feedback was considered self-improvement through adaption. Students within this response category were motivated to address their critical internal feedback by maintaining their original claim and adapting their responses through minor revisions. Students typically proposed and made changes to their responses by incorporating new evidence or reasoning into their responses. Fernald (2 M) does this after engaging in social comparison with the 0 M sample response.

Interviewer: *"What is your confidence in your own response after reading this?"*

Fernald: *"I think that it's a little bit lower because it shows a weakness that I may not have explored in its entirety. And because I don't know the details, I could end up being wrong with my answer."*

Interviewer: *"Okay. What changes would you make after reading this to your answer?"*

Fernald: *"I would probably use, I would ask to see the specific numbers because just guessing kind of off of a graph is not very effective. I'd try to find the ratio that would show that two molarity would be better than zero molarity, unless of course the reverse is true."*

When asked, Fernald began questioning his original response, stating that there was a gap in his response that, if not addressed, could make his response wrong. The potential "weakness" he mentioned had to do with whether or not 2 M had an appropriate amount of gold relative to other metals, something he had relied on with his reasoning in the first stage of the task. Fernald felt there was a gap in his response because he did not include numerical evidence to support his claim. To address the gap, Fernald sought new empirical evidence that would improve and adapt his response. Fernald's response still fit his internal criteria for the task (i.e., amounts of metals in the extraction), but by adapting it through incorporating new numerical evidence he would also definitively align it with his internal criteria.

Self-Improvement: Adoption

Students who found their original response did not fit their internal criteria after the social comparison often made other changes to their response. These changes consisted of adopting a new response with a different claim from their original. This kind of response to critical internal feedback was considered self-improvement through adoption. Students were motivated to wholly address their critical internal feedback surrounding the gaps in their original response and did so by adopting a new claim. Take Hector's social comparison of his 2 M response to the 0 M sample response:

Interviewer: *"How has this affected your thinking about your own response?"*

Hector: *"It kind of made me realize that I didn't account for the single extraction test part. It also enforced that I talked about the gold yield on mine…So it sort of pointed out the things that I liked about mine while also, you know, showing the big point that I ended up missing."*

Interviewer: *"Okay. Is it making you want to change your response at all?"*

Hector: *"Yeah, a bit."*

Interviewer: *"Okay. How would you change your response?"*

Hector: *"If I ended up changing it? I would say I would switch to zero molarity HCl, just because I would want to get out as much gold as I can in a single extraction."*

Although Hector generated some positive internal feedback by identifying a gap in the 0 M sample response that his own response addressed, he generated more critical feedback overall from the social comparison. His original internal criteria were fulfilled by his response, but Hector ended up modifying his internal criteria after the simulated peer review. Hector directly notes that his original response (and his internal criteria) did not account for a single extraction, something that was mentioned as a parameter for the experiment in the prompt. He then incorporated the single extraction criterion into his internal criteria and presumably noted that his original response did not fulfill his full set of criteria. To address this gap, Hector adopted a new claim to better align with the new criteria allowing Hector to better satisfy the updated criteria from the social comparison.

No Internal Feedback Observed

Some of the students in the study did not seem to generate internal feedback from a social comparison. These students stated that they did not have any significant change of confidence after their comparison, and they did not make any changes to their responses. Some of the students in this category mentioned that they did not gain any new information from reviewing the sample response, such as Hal (2 M) did with the 4 M response:

"Just because I guess it would have changed my perspective if I hadn't seen that four molar was the highest extraction. But I already kind of knew that the four molar was the highest extraction going into reading the answer. It didn't really propose anything different or any new information that I hadn't considered." (Hal)

Earlier in his interview, Hal considered 4 M as a choice for his original response, but ultimately decided on 2 M as his final choice for his response. Reading this response exposed Hal to the same evidence and reasoning he had considered before. The lack of new information in the 4 M sample response did little to help generate internal feedback for Hal's own response; therefore, he did not feel motivated to make any changes to his response.

Some students within this category had different reactions to sample responses and would recognize the sample responses as valid. At times, they could even identify what informed that sample response. Consider Jo's (2 M) comparison to the 0 M sample response:

Interviewer: *"Okay. How does this response compare to yours?*

Jo: *"Like I said, different conclusion, most of the same reasoning. Um, I think they're both pretty strong and just have different opinions on the best way to do it."*

Interviewer: *"Okay. What is your confidence in your own response right now?"*

Jo: *"Yeah, it's still the same. I considered all those factors too. I just came to a different conclusion."*

In her response, Jo recognized that her response and the 0 M sample response had similar reasoning, and even considered the 0 M sample response to be a strong argument. She also recognized that its perspective was informed differently than hers, hence the "different opinions" of the responses. Like Hal, she had already considered the information that the 0 M response used and did not feel any differently towards her own response after the social comparison. The social comparison produced no change in confidence and did not seem to offer Jo any critical internal feedback. With the lack of critical internal feedback, students likely did not feel any incentive to revise their responses in any way to improve them.

Discussion

The current study explored how students evaluate their own work in a simulated peer review and what affective changes arose during students' self-evaluations. The reported results advance the community's understanding of the underlying mechanisms of peer review and self-evaluation that can accompany peer review. In doing so, I outlined the cognitive processes students went through to evaluate their own work and identified four distinct outcomes from a simulated peer review. The outcomes identified differ based on how students used internal feedback from a social comparison to evaluate how well their responses met certain internal criteria. The nature of student's generated internal criteria had a deep impact on the social comparisons they made.

To develop their internal criteria, students constructed certain plans to accomplish the task at hand and meet specific goals. To form these goals and plans for any given task, students must consider external information such as the instructor's comments, task prompts, and instructions (Nicol, 2020). Nicol found that the goals that students end up forming to accomplish a task are informed by their prior knowledge, beliefs, experience with similar tasks, and their overall interpretation of instructions given to them. Once students had formed an interpretation of the instructions and constructed goals for their internal criteria, the goals shaped how students evaluated and interacted with all responses for the task, including their own responses. Nicol has also reported that students' criteria formed for a given task influence how they interact with all external products (i.e., their own response and other's responses for a given task) (Nicol, 2020; Nicol et al., 2014).

After producing internal criteria for the task, some students showed evidence of going through the process of decentering. Decentering is the process of recognizing and understanding different perceptions and reasoning from one's own (Piaget, 1955). Decentering has been shown to lead to more productive discourse within the classroom. Physical chemistry students engaging in discourse in a process-oriented guided inquiry classroom demonstrated decentering when they recognized where their peer's response stemmed from, allowing them to consider alternative reasoning and reflect on their own as well (Moon et al., 2017). In this study, students showed evidence of decentering during their social comparison when they could identify the internal criteria that informed the sample response they were reviewing. For example, when comparing the 0 M sample response to his own (2 M) Evander stated that the 0 M sample response weighed the impurities present more heavily in its analysis. Although his own response involved accounting for impurities as well, it was weighed along with the other goal of obtaining a larger amount of gold. This demonstrates that Evander was able to recognize the perspective for the 0 M sample response and identify the internal criteria and reasoning that informed the perspective. It was the act of decentering that allowed some students to develop their analysis and change their responses. Students, such as Hector or Violet, adjusted their own criteria in some way after identifying other internal criteria that informed the sample response they were reviewing.

Students who changed their internal criteria or other aspects of their responses did so because they gathered new information from the social comparison. According to Nicol (2020), students can use comparisons to gather external information to re-evaluate and modify their interpretations of instructions and therefore modify the strategies and tactics they use to accomplish the task. Students participating in this study's simulated peer review had generally engaged in multiple social comparisons before they adjusted their internal criteria or any part of their response. Yan and Brown (2017) also noted this in their investigation regarding student self-assessment. Students used multiple sources of external information to "calibrate" their own performance and evaluations of other's performances. Students generated internal feedback from multiple external sources of information that corroborated each other and then made changes to their work to address the abundance of internal feedback.

Even though students were engaged in multiple social comparisons to generate some sort of internal feedback, there were some instances in which students did not generate any observable internal feedback to evaluate their own work. This can be interpreted as a limitation of the interview setting, as students might have had unconscious internal feedback that was unable to be elicited through the study's interview protocol; however, Nicol's work in internal feedback suggests that providing external information for students to compare against does not guarantee they will make meaningful comparisons to produce internal feedback (Nicol, 2020). Instead, students may "monitor" this external information without using it to evaluate their own work. This "monitoring" could also explain why some students do not make changes to their own work when receiving explicit feedback from reviewers in traditional peer review settings (Finkenstaedt-Quinn et al., 2019). If they are not meaningfully engaging with external information such as peer's constructive criticisms, there is no reason to then generate internal feedback and revise their work.

The students who do end up generating internal feedback and decide to revise their work are likely to be motivated to act in this way. The results suggest that students are acting to address their critical internal feedback and mitigate their uncertainty in meeting their internal criteria as part of this motivation. By addressing their critical internal feedback and working to meet higher standards, students are attempting to improve their work. According to social comparison theory, these students seem to be motivated by self-improvement (Dijkstra et al., 2008). Previous work in peer review has also found that students report that they were motivated to improve the quality of their own work after being exposed to other's work (Nicol et al., 2014). Though this study's model may not capture this motivation wholly as its findings are grounded in a simulated peer review, it is possible that the students in Nicol's study could also be motivated by self-improvement through adoption and adaption.

Conclusions

Findings from this study indicate that when the undergraduate students gave feedback in the simulated peer review, students first formed a set of internal criteria that was used to compare and evaluate their own work against the sample response. Students' comparisons and evaluations served as an opportunity to generate internal feedback surrounding their work. How students' work was meeting their internal criteria for success during their evaluation determined if students generated critical feedback on their work. The nature of students' internal feedback and responses to the feedback shaped four pathways as shown in Table 3.1.

Social comparisons that did not generate critical internal feedback followed one of two response pathways: Self-Enhancement or No Internal Feedback Observed. Students who followed either of these response pathways seemed to lack the incentive to make changes to their work because they did not generate critical feedback. In cases of self-enhancement, students may have even generated internal feedback that validated their work.

In social comparisons that did generate critical internal feedback, students followed one of two response pathways motivated by self-improvement. For either self-improvement pathway, it is likely that the critical internal feedback students generated communicated that students' written work did not meet their internal criteria. In most cases, students' work met most of their internal criteria and only needed new evidence or reasoning to be in full compliance. These students made smaller changes to their written responses to alter it and followed the Self-Improvement: Adapt pathway. In a few cases, students' critical internal feedback communicated that students' written work did not meet any of their internal criteria, and students fully rewrote their responses that adopted a new perspective. Thus, these students followed the Self-Improvement: Adopt pathway.

Table 3.1: Description of student response pathways after engaging in social comparison.

Response Pathway	**Description**
Self- Enhancement	Students who underwent Self-Enhancement displayed higher confidence in their written responses. This likely came from internal feedback that validated their performance in some way. Students then did not revise their work.
No Internal Feedback Observed	Students who followed the No Internal Feedback Observed route did not seem to generate internal feedback, and they did not have any change in their confidence regarding their written responses. They also chose not to revise their work.
Self-Improvement: Adapt	Students who followed the Self-Improvement: Adapt pathway stated that they had lower confidence in their written responses following the comparison. This seems to have been due to their internal feedback telling them that they did not fully meet their internal criteria in some way. Students revised their work but did not change their claims.
Self-Improvement: Adopt	Students who followed the Self-Improvement: Adopt route also stated they had lower confidence in their written work after the comparison. These students likely experienced a large shift in their internal criteria, which meant that their written work no longer fit. Thus, students completely revised their responses and adopted a new claim.

Implications for Research and Practice

The study's investigation of a simulated peer review suggests that peer review settings can help produce internal feedback to help students evaluate their own performance. So long as students engage in a purposeful social comparison with another response, they are likely to generate helpful internal feedback for themselves. Further, the findings show that students generated internal feedback from reviewing both similar and different others. The internal feedback from the comparison can validate the student's work, in which case students will not make changes to improve their performance and maintain their approach. Internal feedback can also provide an incentive for students to revise their performance to improve it. Although our research setting only simulated peer review, the kind of internal feedback observed can be reproduced in actual peer review settings. Many of the findings mirror that of traditional peer review work (Nicol, 2020; Nicol et al., 2014; Nicol & McCallum, 2021), which suggests that the cognitive processes and mechanisms leading to internal feedback from our study are also occurring in traditional peer review settings.

This study's findings, which were grounded in real-time data analysis and interpretation and review, are echoed in other peer review studies that investigated similar processes retrospectively (i.e., focus group interviews following completion of peer review) (Nicol et al., 2014). Follow-up studies need to be conducted to ensure that the processes that were identified occur similarly in an actual peer review setting and that they also lead to the four different outcomes that were observed. Real peer review settings are not always anonymous, nor will students be guaranteed to see a variety of answers such as in this study. Students also tend to receive feedback in traditional peer review, something which I did not include for investigation in this study. In addition to this, the task for the simulated peer review was designed to target two specific performance expectations for the science practice of data analysis and interpretation: analyze data using tools, technologies, and/or models to determine optimal design solutions and analyze data to identify design features or characteristics of the components of a proposed system to optimize it relative to criteria for success (National Research Council, 2012). Future research can and should consider using students' internal feedback to regulate and develop other performance expectations within data analysis and interpretation and consider it for the seven other science practices outlined by the Next Generation Science Standards.

Internal feedback that students generate can also be leveraged in classroom settings. Offering students the opportunity to evaluate preconstructed sample responses allows students to generate internal feedback and evaluate their performance. Findings showed that students were adept at uncovering what ideas and distinctions were contained in a sample response, and how those ideas differed from their own. This means that instructors can leverage preconstructed sample responses to convey ideas, criteria, and nuances in a way that students can understand and use.

Internal feedback can be generated from a comparison with many different external sources of information, so this practice could potentially be expanded to include comparisons with exemplar works or even a rubric for a given task (Nicol, 2020). Facilitating comparisons with sample responses or other sources of external information can be implemented in the chemistry classroom or homework assignments. In chemistry contexts, facilitating comparison can potentially support the teaching of criteria for science practices that have been historically difficult to teach in a lecture or class setting, but are necessary for using chemical knowledge. For example, how to consider all data, weigh variables, and connect data to an assertion are rather difficult features of data analysis and interpretation to teach. This is the case for many of the science practices. Providing more opportunities for students to compare, evaluate, reflect, and revise their own work is a relatively low-labor instructional method that could help to develop certain practices and foster student's own evaluative judgment. These opportunities could serve as a vehicle for having students extract and generate these criteria themselves.

Limitations and Future Opportunities

The findings of this work are unique to the design of the task and the methodological decisions made. Thus, several limitations must be acknowledged.

First, the task itself requires relatively little content knowledge to engage in. This was purposefully done, so that students' revisions would primarily reflect feedback related to the epistemic criteria they used for the task. Tasks that involve the use of more content knowledge could likely elicit internal feedback related to the content knowledge that students use. This would be most apparent in contexts in which students use different content knowledge than the response that they are reviewing. Contexts like these would be more likely to elicit internal feedback and revisions related to the content knowledge involved than the task of this study. Future peer review research should consider how one's content knowledge might influence how one engages in peer review.

Next, I chose to study a simulated peer review that isolated the act of giving feedback to pre-constructed responses. These pre-constructed responses represented anonymous peers to simulate a "blind" peer review that is not representative of all contexts of peer review. Some classroom iterations of peer review may not grant anonymity to the responses that students review, and some students may know the writer of the reviewed response. Knowing the other person might affect how the student evaluates their peer's response, and, in turn, affect how they evaluate their own response. I infer that aspects of Boud's feedback literacy might also affect students' decision to revise, as students may need to manage affect related to comparing to a known person's response and students will need to make judgments about the quality of the known person's response (Carless & Boud, 2018).

Finally, the clinical setting of the think-aloud interview might have affected students' internal feedback generation. Although many students did respond to their internal feedback immediately after reviewing the sample response, it is possible that some students did not have enough time to reflect and respond. Traditional peer review settings often stage the peer review over multiple days, giving students more time to reflect and consider the internal feedback that they generate. The clinical setting of the present study's interviews might have made it so that some students generated internal feedback that they were not aware of while participating in the interview. This opens opportunities for future peer review research to consider ways in which to embed more time for students to reflect on internal feedback.

CHAPTER 4

CHEMISTRY GRADUATE STUDENTS' RESPONSES TO AND SENSEMAKING OF DISCREPANT DATA

Abstract

In preparation for developing into independent scientists, chemistry graduate students must engage with uncertainty to develop new knowledge for their field. One form of uncertainty that graduate students are likely to encounter and must learn how to productively engage with is unexpected data. Considering that graduate students often experience difficulties in developing their independence in their research and that they are likely to encounter unexpected data, there is a need to better understand how students engage with unexpected data. Thus, this study seeks to investigate how chemistry graduate students respond to unexpected data and how these responses affect their sensemaking of a chemical phenomenon. Chemistry graduate students participated in think-aloud interviews in which they analyzed multiple data sets to explain a chemical phenomenon. Using these interviews, I identified instances wherein students encountered data that was discrepant with their predictions and used Data-Frame Theory to characterize how students responded to discrepant data. Additionally, students' sensemaking that followed was analyzed to identify how the response affected students' identification of underlying causes for the phenomenon and establishment of connections between the data sets. Graduate students showed three distinct responses to discrepant data: preserving the frame, elaborating the frame, and reframing. All three responses were capable of progressing students towards productive sensemaking of the phenomenon, but students needed to use the response to accomplish the goals of the analysis. Implications for supporting graduate students' analyses of discrepant data are also discussed.

Introduction

Chemistry graduate programs are meant to train graduate students to be competent and independent researchers who contribute new knowledge to their field through original research (Donkor & Harshman, 2023; National Academies of Sciences, 2018; The American Chemical Society, 2012). Knowledge is advanced within science by engaging with the uncertainty of the world (Kampourakis & McCain, 2020). One form of uncertainty that graduate students are likely to encounter in their research is data that is discrepant with one's expectations (Grolemund & Wickham, 2014). Thus, if chemistry graduate students are to be trained to contribute original research that advances the knowledge of the field, they must also be trained to productively engage with unexpected data.

Of the chemistry education studies that have considered chemistry graduate students' data analyses, both studies have investigated how chemistry graduate students analyze sets of IR and ^{1}H NMR data to characterize molecules. Cartrette and Bodner interviewed fifteen chemistry doctoral students as they analyzed IR and ^{1}H NMR data to construct structures for different organic molecules (2010). The authors identified that students who successfully determined the molecular structure checked the proposed molecular structure against the spectra to ensure the data supported their claims. Additionally, more successful students "mined" the spectral data to find important information, such as coupling constants in the ^{1}H NMR, to aid in proposing a molecular structure. Similarly, Connor and colleagues tasked chemistry graduate students to analyze IR and ^{1}H spectra to determine if an organic molecule was synthesized or not (2021). Using eye-tracking, the authors found that doctoral students spent less time looking at less relevant information when determining if a molecule was synthesized. The authors also identified that graduate students searched for complementary data across the IR and NMR spectra, which suggests that they looked for corroborative evidence to support their claims across the different data sets.

Although both studies offer important information on chemistry graduate students' analyses of spectral information, neither study considers what students do when encountering unexpected data. Given that many doctoral students face challenges in their development as independent scholars (Gardner, 2007, 2008), research needs to explore this facet of chemistry students' data analysis and identify ways in which to support their development. Specifically, an exploratory qualitative study is needed to investigate graduate students' sensemaking of discrepant data. An exploratory qualitative study can describe a rather complicated phenomenon (sensemaking of discrepant data in this case) that cannot be easily measured with existing statistical instruments like surveys (Creswell & Poth, 2016). To do this, interviews that directly elicit students' reasoning are needed to generate rich, detailed accounts of students' sensemaking (Creswell & Poth, 2016; Merriam & Tisdell, 2016). These accounts can be analyzed in such a way to develop a model that can describe students' responses to discrepant data and explain how students' responses impact their sensemaking (Creswell & Poth, 2016; Merriam & Tisdell, 2016). Thus, qualitative research is the most appropriate form of inquiry to address this research gap.

The following study explores how chemistry graduate students respond to and make sense of data that was discrepant with their expectations. The graduate students in the following study participated in think-aloud interviews as they made sense of multiple data sets to explain a chemical phenomenon. Data-Frame Theory was then used to characterize students' responses to discrepant data. I also explore how students' responses affected the progression of their sensemaking and explanation construction. With these in mind, this study was guided by the following research questions:

1. How do chemistry graduate students respond to discrepant data when analyzing multiple data sets to explain a chemical phenomenon?
2. How do the different responses to discrepant data differentially affect sensemaking?

Theoretical Framework

Klein and colleagues' Data-Frame Theory was used, as it characterizes how one's reasoning can develop throughout one's sensemaking of data (Klein et al., 2007; Klein & Moon, 2006). It can explicitly outline the ways one's reasoning interacts with one's observations of data through its characterization of different sensemaking activities. This is especially important to consider for this research context, as the framework's defined sensemaking activities can be used to characterize how graduate students respond to discrepant data.

To begin, Data-Frame Theory posits that there are two interrelated components involved in one's sensemaking. The two components, the data (observations) and frame (reasoning), are constructed simultaneously throughout the sensemaking process and interact in distinct ways. Data can be considered any directly observable feature of the thing one is making sense of. For the context of data analysis, data can take the form of tabulated measurements or visual representations such as graphs. Data includes any directly observable feature of the data set, including individual points, trends, patterns, peaks, slopes, data labels, legends, etc. The other component of Data-Frame Theory, the frame, is an explanatory structure that serves to ascribe meaning to observations from data by describing how data relates together. Frames come in many forms, including scripts, plans, prior knowledge and experiences, and schemas (Berg &

Moon, 2023; Gouvea et al., 2019; Grolemund & Wickham, 2014; Hammer et al., 2004; Klein et al., 2007; Klein & Moon, 2006; Zhou & Moon, 2023).

To begin the sensemaking process, one makes initial observations of the data. These observations activate a frame, which prompts one to recall prior knowledge and experiences, schemas, and other cognitive components. Once some aspect of a frame has formed, the frame begins to “filter out” data that is less relevant and identify more important information to consider. The incoming information may also elicit more pieces of the frame to appear, and the new parts of the frame may continue to filter out and search for more data. This cycle of observing data and forming one’s frame iterates until the frame and data converge.

During the sensemaking process, there are times in which the data observed will not fully align or fit with one’s frame (Grolemund & Wickham, 2014; Klein et al., 2007). I have defined these moments as discrepancies. When this occurs, the analyst might question their frame and notice where inconsistencies exist between the data and frame. This is akin to problematization, as the analyst identifies that there is a gap in their frame such that they cannot explain the data (Phillips et al., 2018).

Often, the analyst will elaborate their frame in response to encountering unexpected data. That is, they maintain the core tenets of their frame and add new information to better integrate the data with the frame. This may involve searching for more information within the data itself and going through data that may have been overlooked or misread previously. Elaborating the frame can also involve incorporating additional conceptual knowledge or drawing upon previous experiences to help make sense of the discrepant data. This new information is then added onto the frame to account for the unexpected data, but it does not challenge or call into question any initial components of the frame.

In the case that discrepant data cannot be integrated into the frame through elaboration, the analyst may prioritize fully preserving their frame. When this occurs, the analyst does not consider any additional information in their frame, and they may ignore, discount, or disregard the discrepant data entirely. Essentially, the frame stays intact, and the discrepant data can be "filtered out" so that sensemaking can progress. This response has been commonly observed in studies regarding conflicting or anomalous data, in which students ignore, discount, or explain away data that does not fit their mental model (Bolger et al., 2021; Chinn & Brewer, 1998; Meister et al., 2021; Urbanek et al., 2023). However, there may be instances where the analyst may have to consider altering their frame to account for the discrepant data, especially in cases where the discrepant data does not support one's conclusions.

When the unexpected data is accepted but cannot be integrated into the frame through elaboration, the analyst likely needs to engage in reframing. To do this, the analyst must disengage from their current frame. This means that the core ideas composing the frame are no longer used by the analyst, so the analyst may construct another frame that can explain the discrepant data. In constructing their new frame, the analyst will identify new core ideas that help to make sense of the data. Reframing is similar in some respects to Chinn and Brewer's "theory change" in which analysts dismiss previous conceptions to account for anomalous data (1998).

When engaging in sensemaking of data, analysts are likely to use any combination of preserving, elaborating, and reframing their reasoning. These sensemaking activities all serve different functions to help one's analysis progress depending on the nature of the discrepancy between the data and frame. Thus, it is likely that an analyst will employ different activities at different points in their sensemaking. This is especially true for making sense of multiple data sets, in that there may be gaps between data sets that require additional content knowledge to fill, extraneous data that must be filtered out, and predictions across data sets that may not carry over. Analysts employ these different activities iteratively until their frame and the data converge or the analysts consciously choose to stop.

Methods

Task Design

I first designed a task that used multiple data sets that were accessible to different chemistry graduate students. The task was designed using data from a recently published physical chemistry paper (Bodesheim et al., 2020), as it featured multiple data sets in the form of graphs. Additionally, the data sets involved measurements related to intermolecular forces and thermodynamics, which are foundational concepts that appear across the different sub-disciplines of chemistry.

In the paper, Bodesheim and colleagues seek to explain why hydrogen fluoride (HF) does not change phase from an orthorhombic (36) to cubic (225) structure in the same temperature range as three other hydrogen-halides: hydrogen chloride (HCl), hydrogen bromide (HBr), and hydrogen iodide (HI). To do this, the authors performed a series of computations that compare the energies of the hydrogen-halides in the 36 and 225-structures. Three of the resulting computational data sets featured in the paper were chosen to be used for the task.

These data sets were used for the task, as they could be scaffolded to be accessible for chemistry graduate students to analyze. Additionally, I predicted that using and establishing connections between multiple data sets might lead to moments in which students might get “stuck” or encounter data they did not expect.

Interview Protocol Design

Because of the complexity of the task, teaching interviews were used to surface graduate students' sensemaking. Teaching interviews have been used previously to capture student sensemaking and abstraction of information (Adams, 2023; Hershkowitz et al., 2001; Kapon & diSessa, 2012; Karch & Sevian, 2022). In teaching interviews, the interviewer provides some level of assistance to the participant. This assistance can take the form of explaining concepts to students or scaffolding connections that participants can make but may not immediately notice when engaging in the task. Ultimately, the goal of using the teaching interview format was to provide enough information and background so that students could more deeply engage in sensemaking even though they were unfamiliar with the context.

The teaching interview took place in stages so that students would be introduced to one data set at a time before they engaged in constructing an explanation. Each data representation was scaffolded so that students had relevant background information to make sense of the data. This information included basic experimental details, some mathematical relationships (i.e., equations), and mathematical information (e.g., the difference in conformational entropy is negative) to help students connect the vibrational entropy to other thermodynamic values. The interviewer also repeated background experimental information as needed to the students and would point out data features (such as labels and axes) if students seemed to overlook them.

The interview began by introducing students to the two structures involved in the phase change (the 36-state and 225-state) of the hydrogen-halides using figures from the publication (shown in Fig. 4.1). After students made observations of the structures, students considered experimental data that illustrates the central phenomenon, that HF did not change to the 225-phase like the other hydrogen-halides.

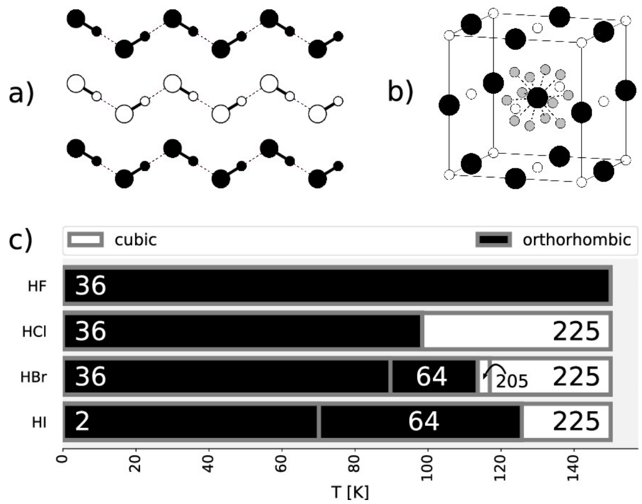

Figure 4. 1: Crystal structure and phase chart figure modified from Bodesheim *et al.* (2020). (a) The chain structure of the 36-structure. The larger atom represents the halogen, and the smaller atom represents the hydrogen. The black shading indicates the rows existing in the front plane, and the white shading indicates the rows existing in the back plane. The dashed line represents the hydrogen bonding between molecules. (b) The crystal structure of the 225-structure. The large, black atoms represent the halogens, and the small, white atoms represent the hydrogens. (c) A summary of the different hydrogen-halides reported to date of publication. The numbers represent the space groups of the crystal structure.

Next, students analyzed the intermolecular force bar graph, Figure 4.2, in which they were prompted to compare the energies of the different hydrogen-halide structures' intermolecular forces. In this stage, the graduate students were told that the data set distinguished the intermolecular forces for each hydrogen-halide structure by its van der Waals (vdW) and non-van der Waals (non-vdW) interactions. All participants were also told that van der Waals forces could be thought of as London dispersion forces in this context.

Participants then analyzed the thermodynamic line graphs shown in Figure 4.3. During this part of the interview, students were given equations that showed that values plotted in the graph were composed of a difference between the two structures' energies. The graphs depicted the difference in Gibbs free energy, the enthalpy, and entropy between the two structures, specifically the 36's value minus the 225's value. Questions in this stage targeted what differences students observed between the hydrogen-halides' line graphs, what relationships students identified between the thermodynamic values, and how the thermodynamic data related to the phase change.

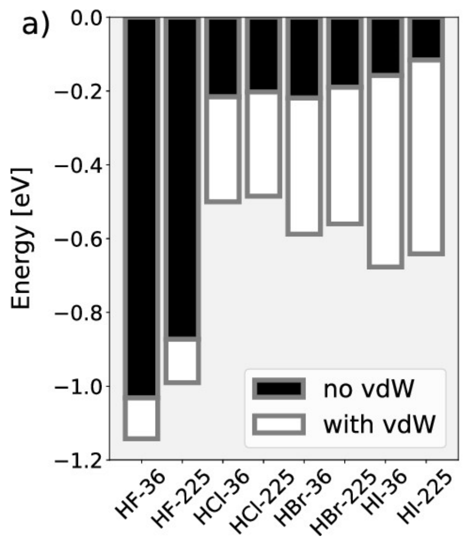

Figure 4. 2: Intermolecular force energy bar graph modified from Bodesheim *et al.* (2020). Graph depicts energies associated with HX structure's intermolecular forces. The black portion represents the intermolecular force energy without the vdW correction, the black and white portion represents the intermolecular force energy with the vdW correction.

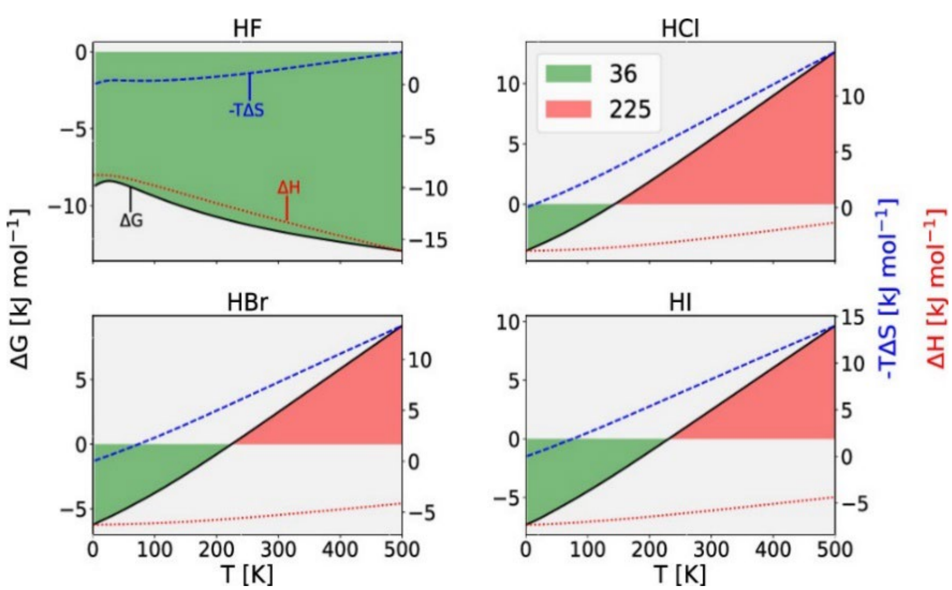

Figure 4. 3: Thermodynamic line graphs depicting differences in Gibbs free energy, enthalpy, and entropy from Bodesheim *et al.* (2020). The solid black line represents the difference in Gibbs free energy, specifically $\Delta G = G_{36} - G_{225}$. The dotted red line represents the difference in enthalpy, specifically $\Delta H = H_{36} - H_{225}$. The dashed blue line represents the difference in -TS, specifically $\Delta S = S_{36} - S_{225}$.

Next, the graduate students were introduced to the background information surrounding the entropy, wherein participants were specifically told that the total entropy had two contributors. In this, the interviewer told the participants that all hydrogen-halides had an equivalent conformational entropy difference and that the conformational entropy difference between the 36- and 225-structure was negative, meaning that the conformational entropy favored the 225-structure. Students then thought aloud to reason why HF had a different entropy than the other hydrogen-halides.

In the final data set, students were introduced to the concept of vibrational entropy (that it measures the space explored through atoms' vibrations (Fultz, 2010)), given some basic experimental background information, and directed to the relevant data features before they began to analyze the vibrational entropy data shown in Figure 4.4. The interviewer also asked

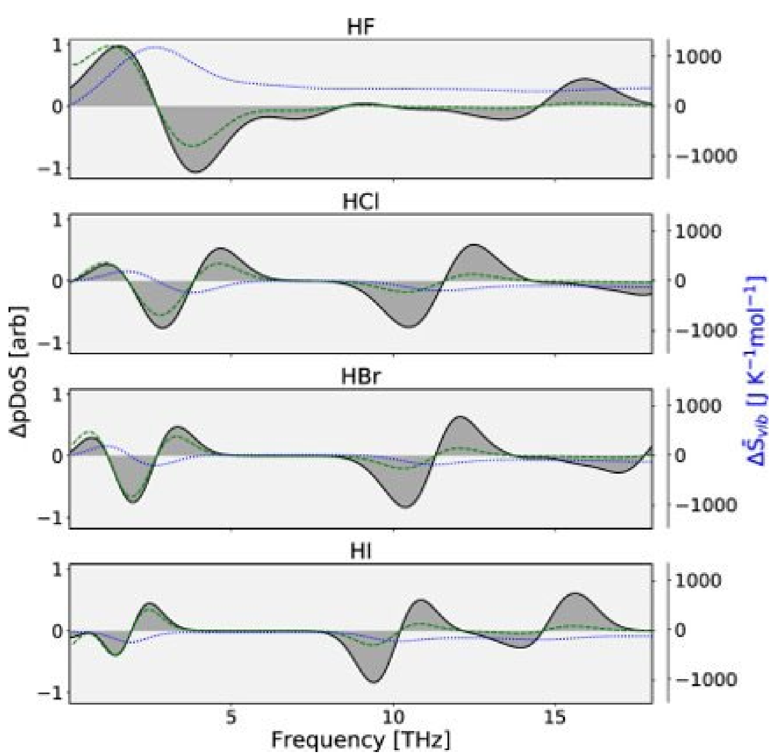

Figure 4. 4: Difference in pDoS figure modified from Bodesheim *et al.* (2020). Students were directed to only consider the blue line, which represents the integrated difference in vibrational entropy or $\Delta\bar{S}_{vib}$. Data is calculated at 300 K.

questions meant to prompt students to mathematically relate the vibrational entropy values in this data set to the previous thermodynamic data.

For the last stage of the interview, participants were prompted to explain why HF did not change phase like the other hydrogen-halides using the data they thought was necessary for an explanation. Participants were encouraged to review previous data sets and take notes as needed. This final stage of the interview was driven by the participants reasoning aloud, but additional questions were asked to probe students' reasoning and ask students how they thought the different data sets were related or connected. In every interview, students were asked how the vibrational entropy data related to the intermolecular force data if the students had not brought it up themselves. Once the participants stopped bringing up new ideas or voiced that they needed to finish, the interview ended.

For the full draft of the interview protocol and explanation given to participants after the interview was complete, see Appendix B.

Participants

All participants included in the study were graduate students currently enrolled in a chemistry doctoral program at the time of the study. Institutional Review Board approval was obtained before interview recruitment was started. The participating graduate students were recruited from multiple research-intensive universities in the Southeast, Midwest, Southwest, and Pacific Northwest regions of the United States. I purposefully sampled interested students during recruitment so that participating graduate students (N=23) ranged in experience from students in their first year to students in their sixth year of study in their doctoral program (Creswell & Poth, 2016). I also purposefully sampled so that participating students represented multiple sub-disciplines of chemistry, including materials chemistry, polymer chemistry, laser chemistry, organic synthesis, inorganic chemistry, analytical chemistry, chemical biology, and computational chemistry. Before each interview, participants completed and submitted electronic consent forms. After their interview, participants were given a $20 digital gift card as compensation for their time.

Data Collection

All interviews were conducted from Fall 2022 to early Spring 2023. The interviews were conducted using Zoom and ranged in length from thirty-five to ninety minutes. All interviews were recorded within the application and the resulting video recordings were used to generate transcripts of the interview. The video recordings were also used for analysis, both when participants annotated the data representations on their computer screens and when participants made gestures with their hands in their cameras. All transcripts and videos were deidentified and participants were assigned pseudonyms. Any notes participants wrote during their interviews were also collected as an additional interview artifact for analysis.

Episode Selection

To find moments in which participants encountered discrepant data, I used verbal cues in which participants explicitly mentioned that they expected to observe something different in the data, a data feature did not align with their expectations, or something directly conflicted with their reasoning. Each episode started at the point in which the participant identified the discrepancy, meaning they explained what they expected to observe and described what they actually observed in the data. The episode ended when the participant seemed to stop their sensemaking of the discrepancy. This point was identified as when the participant stopped bringing up new ideas in response to explaining the discrepant data (i.e., their explanation was exhausted) or alternatively when the participant stopped discussing the discrepant data and moved on to analyzing other data features.

I used two criteria to select episodes before beginning the analysis. I first sorted out episodes in which the participants did not clearly specify what it was that they were expecting to observe in the data. I also sorted out the episodes in which the participants' sensemaking of the unexpected data was interrupted and the participant did not bring up the discrepancy again.

In the remaining collection, there were instances in which some participants seemed to resurface the same discrepancy across multiple episodes. Because these episodes involved the same discrepancy, they were collapsed into single episodes. Using these criteria for selecting, sorting, and collapsing episodes, I identified thirty-seven episodes that occurred across twenty-three participants' interviews (N=23).

Data Analysis

The data analysis took part in three stages. First, I identified what may have caused the student to view the data as discrepant or what made the data unexpected. Next, students' responses to the discrepant data were characterized using Data-Frame Theory. Finally, I explored how students' responses to the data affected their following sensemaking and explanation construction.

First, the analysis sought to identify what led students to perceive the data as unexpected. To do this, Data-Frame theory was used to help identify what Data-Frame components were involved. For each episode, I first identified what students expected to see in the data versus what it was that they observed. In describing students' expectations, I identified the expectations resulting from specific components of students' frames (i.e., what conceptual knowledge or ideas were used) and/or the expectations resulting from other data features (e.g., patterns). For example, consider Ashland as they attempted to relate their interpretations of the intermolecular force data and the vibrational entropy data:

Ashland: *"So undergoing the [225 to 36] transition increases the vibrational entropy, meaning that [the HF molecules are] like having a greater area to explore [in the 36 phase], or the vibrations can spread out across the larger area... Which I wouldn't have actually expected at gut instinct."*

Interviewer: *"Yeah, what was your gut instinct?"*

Ashland: *"It's that if you're having really strong, electrostatic, or intermolecular forces, that [molecules] are going to be held more rigidly in place because they're going -- or their motion is going to be more hindered by their interactions with other molecules. Is what I would have expected."*

Here, Ashland identified that their interpretations of the intermolecular force data and the vibrational entropy data conflicted. Specifically, they noted that they did not expect HF 36 to vibrate across a larger area when HF had strong intermolecular forces that "held" molecules rigidly in place. Because this reasoning of the intermolecular force data is not something that can be directly observed in the data, their reasoning seems to be informed by a knowledge component of their frame, "stronger interactions 'hold' molecules in place." This reasoning conflicted with their interpretation of the vibrational entropy data. Ashland had observed that the transition from 225 to 36 produced an increase in vibrational entropy, and they interpreted that to mean that HF molecules would explore more space via vibrations in the 36-phase. Ashland's interpretation was likely guided by another knowledge component of their frame, "more vibrational entropy means more space explored." This knowledge component of their frame that dictates HF molecules are moving conflicts with Ashland's other part of their frame that says HF molecules are "held" in place. Thus, Ashland's discrepancy would involve these conflicting knowledge components of their frame.

Next, to identify how students responded to the discrepant data encounter, I process coded a subset of the episodes (Miles et al., 2014). This involved describing the actions that students did throughout their sensemaking, such as identifying the discrepant data, making observations of data features, etc. The process codes from these episodes were used to develop an initial codebook. I then applied this codebook to the remaining episodes, revising to capture new actions until all actions students used to make sense of data could be accounted for.

Then, summaries of each episode were constructed that began with students' identification of the discrepancy and ended with the last action that students used. Using these summaries, I examined how students' sensemaking actions affected their original line of reasoning (i.e., did students alter their initial reasoning in any way). Based on this step, I separated students whose sensemaking moves did not alter their original reasoning from students whose reasoning did change in some way. I categorized the episodes in which students did not alter their reasoning as "Preserving frame," as their reasoning did not change. Students in this group were further differentiated based on what moves were used after they encountered the discrepancy, which resulted in two sub-categories of preserving frame sensemaking moves: students dismissing data and students not using new reasoning. For students who did alter their reasoning, I decided to further separate based on the extent of the change in their reasoning, which resulted in two categories. In one group, students altered their frame by expanding on their original reasoning. This expansion could involve using mathematical relationships, incorporating new conceptual knowledge, or accounting for new data features. Thus, I labeled this category "Elaborating frame" as they added onto their frame in some way and did not replace any original ideas. For the other category, students dismissed or stopped using an idea that was involved in their discrepancy and replaced it with a new frame component. This last category was labeled "Reframing" as these students uniquely dismissed and replaced an essential frame component involved in their discrepancy.

Finally, to explore how students' responses affected the sensemaking that followed, I considered how the sensemaking fulfilled the purpose of the task. Students were told that the main purpose of the task was to construct an explanation as to why HF did not change phase like the other hydrogen-halides. In doing this, a student's sensemaking was deemed most productive when they "connected the dots" between data sets or recognized the relationships between the different data sets (Klein et al., 2007). Additionally, students' sensemaking should have identified the causes for what they observed in the data, both why HF differed from other hydrogen-halides and why one structure was preferred over another. These points were used as criteria to assess if the response contributed towards more productive sensemaking or limited sensemaking. If more criteria were fulfilled (i.e., connections were made, underlying causes identified, etc.), students' sensemaking was considered to be more productive. Conversely, less productive or limited sensemaking occurred when the students' explanations were missing these criteria.

Trustworthiness

Due to the exploratory nature of this work, I utilized trustworthiness as a method to establish the credibility and dependability of the analyses (Lincoln & Guba, 1985). To do this, an outside researcher familiar with the task was trained on the codebook. Additionally, the outside researcher was trained in how to identify the Data-Frame components involved in students' discrepancies.

The outside researcher and I then independently analyzed eleven episodes to identify what may have caused students to perceive the data as unexpected. The outside researcher and I met to compare and discuss the selected episodes until consensus was reached. The discussion centered on when students were using patterns within the data to create expectations for other data (e.g., other hydrogen-halides' lines increased when HF's line did not) and when students used content knowledge to create expectations. We decided that students needed to explicitly voice if a data feature was not following a pattern that the other data features did for their expectations or predictions to be shaped by patterns in the data. We also decided that students needed to explicitly voice what of their content knowledge was not aligning with the data (such as energy should not increase with temperature) to account for students' frames to be shaped by their frame.

Next, the outside researcher and I independently coded the eleven episodes to identify how students responded to the discrepant data and decide if this response seemed to contribute toward more productive sensemaking. We then met again to compare and discuss the coded subset of episodes until we had reached consensus. The main point of discussion involved disagreement on whether certain episodes showed students using one response or another. For example, in discussing one graduate student's response (specifically Gibbon), the researchers disagreed if the student preserved their frame or elaborated their frame in response to a discrepant data feature. We came to an agreement that the student responded to the discrepant data feature by preserving their frame; however, following this response, the student then mathematically connected the data set they were analyzing to another, which was indicative of elaborating their frame. Because the student preserved their frame before they had elaborated their frame and because the frame preservation occurred in response to the discrepant data feature rather than the entire data set, the researchers decided to classify the student's response as preserving frame. It is very common for analysts to use multiple sensemaking activities (i.e., preserving frame, elaborating frame, etc.) when making sense of data though (Klein et al., 2007), so the student's frame elaboration was considered a part of the sensemaking that followed the response to the discrepant data.

Following these discussions, I updated the codebook to provide clearer descriptions of students' discrepancies and applied updated codes accordingly to the remaining data.

Results

Overview

Graduate students appeared to respond to unexpected data in three different ways: preserving the frame, elaborating the frame, and reframing.

I first explore students preserving their frame in response to their discrepancy. Students who preserved their frames did not alter their reasoning and did not attempt to incorporate the discrepant data into their frames. This worked to preserve students' original state of mind.

Next, I examine students' responses to their discrepancy by elaborating their frame. Here, students built upon their original reasoning to account for the discrepant data. However, students did not replace their original reasoning or original observations of the data.

Finally, I investigate students reframing after encountering their discrepancies. Reframing involved students dismissing some aspect of their original reasoning to accommodate the discrepant data. This helped students to start afresh and build new reasoning that could more closely align with the discrepant data.

The findings are presented in greater detail below with examples that feature quotes from students' interviews. Each student example also features two figures. One figure describes students' discrepancies and shows the data and reasoning involved. The second figure summarizes how students' responses to the discrepancy affected their sensemaking.

Preserving the Frame

One way in which graduate students responded to discrepant data was to preserve their frame. When students preserved their frame, they did not offer new ideas or account for the discrepant data in their explanation. Thus, students did not alter their reasoning to incorporate the discrepant data into their explanation, and there may still have been a gap in students' understanding of the discrepant data.

There were several ways in which students preserved their frame in response to the discrepancy. After students identified the inconsistency between their expectations and the data, some students proceeded to move on to analyzing other data features immediately, which suggests they might have ignored the discrepancy. Conversely, some students spent longer periods reiterating what they had expected to see in the data, and they did not attempt to alter their reasoning to the discrepant data. Lastly, a small subset of students discounted the discrepant data feature to help them move forward in their analysis of the data. Consider Fremont's reasoning that led to their discrepancy and their response:

Before Fremont's discrepancy encounter, they used evidence from the vibrational entropy data set to establish why HF 36 would not spontaneously form by connecting it to the Gibbs energy data.

"So if we look back at here, that explains why you've got this ΔG dropping as the temperature increases, and the ΔG increasing for the other three [hydrogen-halides]. So when I take all that into account, I'm thinking because the vibrational entropy is so large in magnitude for the HF [36] if we're looking at this [vibrational entropy data] slide and because it's always

positive, that alone is going to completely almost eliminate the possibility of it, wanting to have the proper energy value to spontaneously form the cubic. Whereas because these other [hydrogen-halides'] have negative [vibrational entropy] values, when you plug all of those back into the overall equation that then becomes a favorable [Gibbs] energy value to spontaneously form the cubic where HF can't. At least that's how I interpreted it." (Fremont)

At this point, Fremont's explanation construction established a cause for HF not changing phase like the other hydrogen-halides; the vibrational entropy did not favor a phase change. However, their explanation did not establish a cause for why the vibrational entropy did not favor a phase change, nor did it connect the intermolecular force data. Thus, Fremont was prompted to consider how the intermolecular force data affected the thermodynamic data.

In response to this question, Fremont fixated on the HF 225 bar present in the intermolecular force data and pointed out that it was discrepant with their expectations from the previous analysis. Specifically, Fremont focused on the magnitude and value of the HF 225 bar, reasoning that because the HF 225 bar had a relatively similar magnitude to the HF 36 bar, the HF 225-structure must be present in the same relative abundance as HF 36.

Fremont: "*Okay. So yeah, this completely negates the 225 column for HF because if we were seeing this, if it was this preference like preferential, you'd expect a lot of the [HF] 225 to form. Because if you're trying to, if you lose, energy is dropping. That means it's more favored in the state it's in, correct?"*

Interviewer: "*Mhm."*

Fremont: "*Yeah, so it should be a lot more, [HF 225] should be like almost equally favored. I mean not quite equal, but more or less equal to the orthorhombic. But that is not what we're seeing at all in the previous slide or by calculations. Like it shouldn't want to form the HF*

225 the cubic form, but just based off of the initial computational calculations it looks like they should."

Here, Fremont seems to be reasoning that because the HF 225 bar is most similar in magnitude to the HF 36 bar and because it has such a negative value, the HF 225 should preferentially form. Fremont's interpretation of the intermolecular force data here seems to have clashed with their other interpretations of the vibrational entropy and thermodynamic data sets, in that HF 225 should not have formed whatsoever.

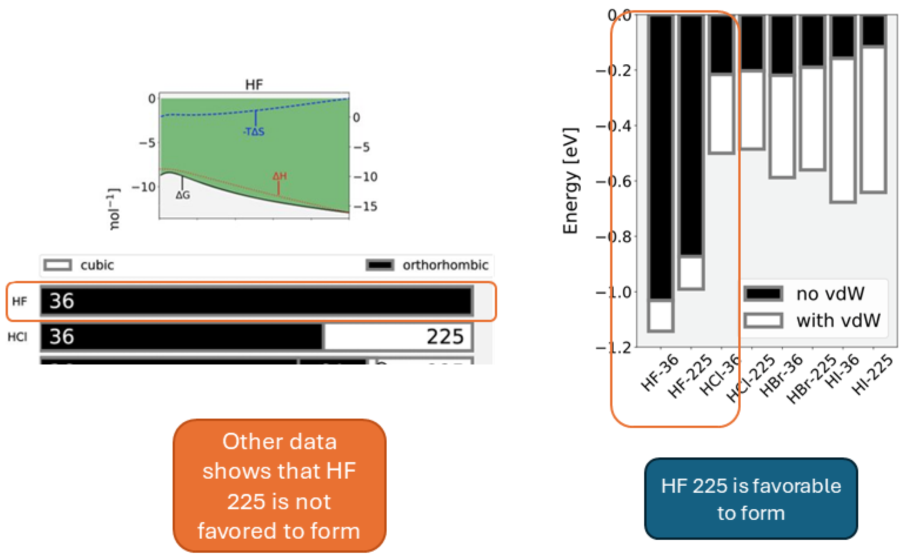

Figure 4. 5: Fremont's discrepancy involving data from Bodesheim *et al.* (2020). Fremont original reasoning from the experimental phase chart and thermodynamic data that concludes HF 225 should not form (orange). Fremont's original reasoning from the intermolecular force data that HF 225 is favorable to form (blue).

In response to this discrepancy between the intermolecular force interpretation and their other interpretations, Fremont dismissed the HF 225 bar in the intermolecular force data.

Fremont: *"So the experimental data kind of eliminates this [HF 225] column entirely, and it should not be there. I mean, it's there, but it's not going to be in the real-world application . . . [The experimental data] kind of negates that previous [intermolecular force] calculation, I think.*

Interviewer: *"Okay. Can you remind me what computation you say was being negated?"*

Fremont: *"So you said this was all computational data. Like they didn't actually verify this, this was just theoretical. So based off whatever they used to calculate this, it should be almost like fairly equally favored for the HF 36 and the HF 225 to form. Like not quite as close as the others, but they're still within, like reasonable, like intensities and magnitude of each other. But then, if you go on to the actual, like looking at the individual [thermodynamic] components of the data for like ΔG, ΔH, TΔS, you don't see that distribution at all. Like [HF] 225 never shows up. So the, if you're looking at this [intermolecular force] calculation, this [data] is an outlier."*

After explaining why this data feature should be dismissed, Fremont stopped their analysis and ended the interview. This prevented them from identifying a deeper cause for the thermodynamic and vibrational entropy data (i.e., why HF 36 had more vibrational entropy), which can only be explained by using the intermolecular force data. Specifically, Fremont did not consider the proportions of van der Waal and non-van der Waal contributions in the intermolecular force data, and they also did not seem to consider the difference in intermolecular force energy between the HF 225 and HF 36 structures either. These details were important in establishing causes for why HF favored HF the 36 structure vibrationally and consequently, why HF 36 was lower in Gibbs free energy than HF 225. Through dismissing the discrepant intermolecular force data feature, Fremont seemed to stop analyzing the intermolecular force data set entirely. This prevented them from connecting the intermolecular force data to other data

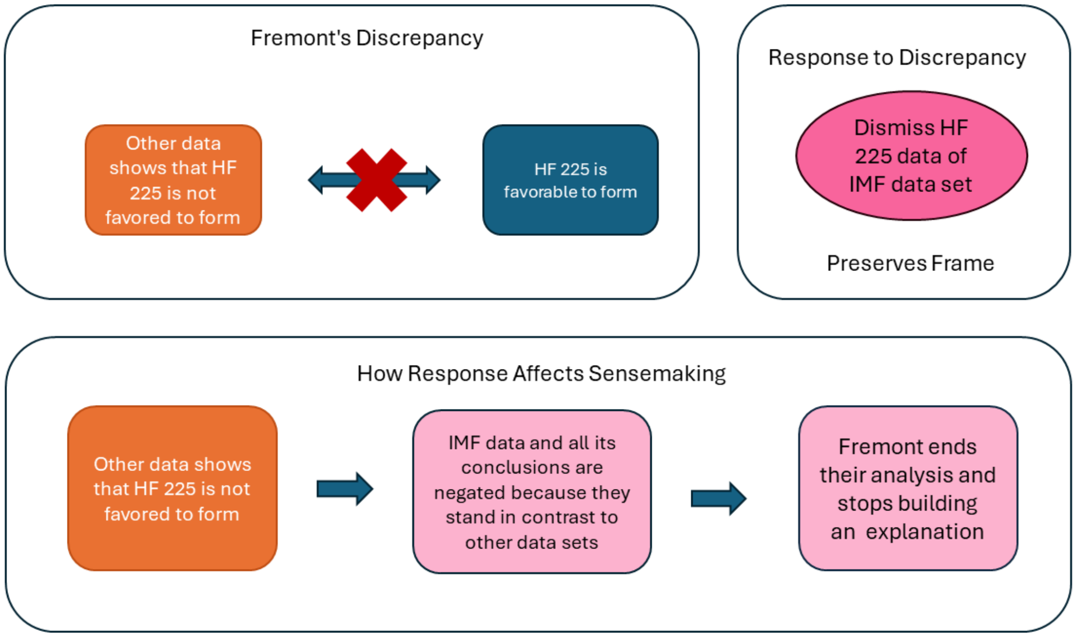

Figure 4. 6: Graphic illustrating Fremont's discrepancy, response, and sensemaking. Upper left: Fremont's discrepancy between their reasoning from both the other data sets (orange) and the intermolecular force data set (blue). Upper right: Fremont preserved their frame (darker pink). Bottom half: Fremont preserved their original reasoning (orange) by using reasoning that dismissed the discrepant data (lighter pink).

sets and from identifying any further causes for the chemical phenomenon. Thus, this response limited some of the sensemaking Fremont could engage in to explain why HF did not change phase.

In some instances, preserving one's frame could contribute toward more productive sensemaking. When students encountered data that they did not expect, it took time and cognitive effort to make sense of the data. In some instances, it may have been important to engage in further analysis of the discrepant data feature if, for example, it helped students to fulfill the goals of the analysis (i.e., establish connections, identify underlying causes), such as with Fremont. However, there were some contexts in which the discrepant data feature did not help students' sensemaking accomplish the goals of their analysis. In these times, it was more appropriate for students to preserve their reasoning and move ahead to analyzing other data features.

For example, Gibbon attempted to relate the vibrational entropy values of the different hydrogen-halides to their total entropy values in the thermodynamic data. During this time, Gibbon noticed that HI's total entropy in the thermodynamic data behaved very similarly to HCl and HBr's total entropies. However, when considering the vibrational data, Gibbon noticed that at lower frequencies, HI's vibrational entropy behaved differently from that of HCl and HBr; from around 0 THz to 5 THz, HI's vibrational entropy stayed negative or was at zero, whereas HCl and HBr's vibrational entropy had peaks with positive values. Gibbon did not seem to expect HI's vibrational entropy to differ from the two other hydrogen-halides and voiced that the two data sets misaligned:

"And so that [the total entropy data] doesn't really line up with what I was talking about here [in the vibrational entropy data], where HI kind of looks different from the HCl and HBr. I guess that could mean that the [vibrational entropy] contributions from the really low frequency

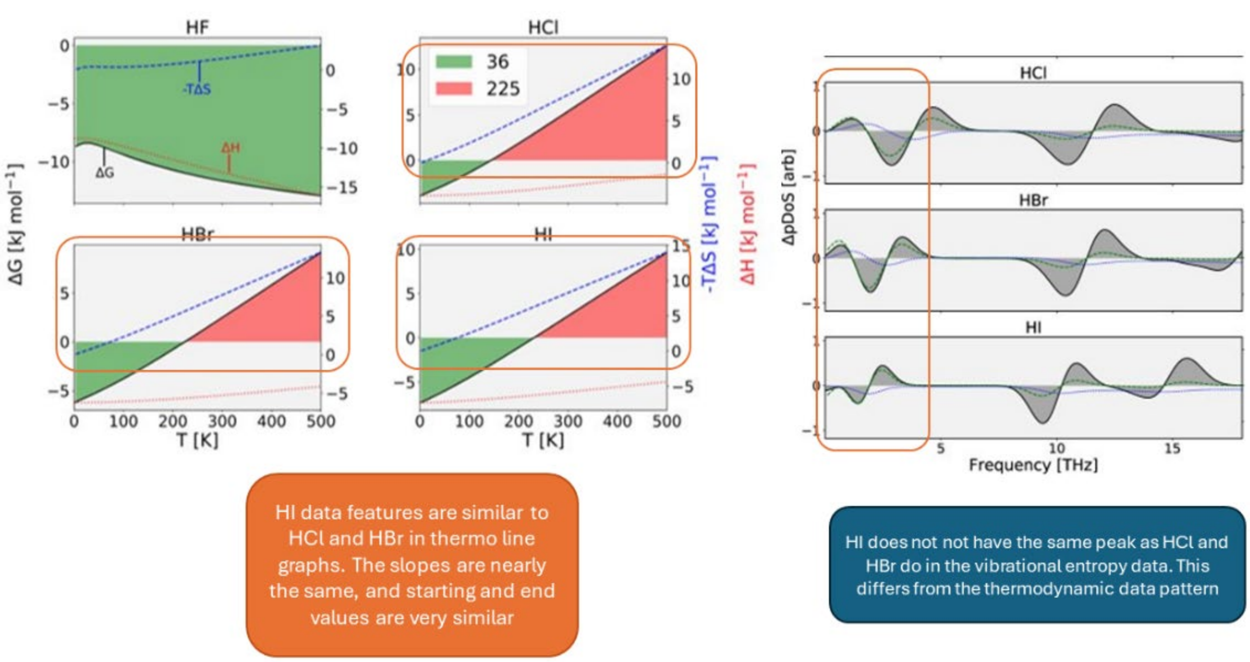

Figure 4. 7: Gibbon's discrepancy involving thermodynamic line graph and vibrational entropy data from Bodesheim *et al.* (2020). Gibbon noticed that a pattern between HCl, HBr, and HI entropy lines in the thermodynamic line graphs (orange) was not replicated in the vibrational entropy data set (blue).

are less important. So maybe the overall [vibrational entropy] term is negative, and then these little components ****moves cursor over first HCl bump in vibrational entropy chart**** *maybe don't matter as much. And so that could be why the, like overall the [total entropy] sum looks really the same for all three."* (Gibbon)

After Gibbon noticed that HI's resemblance to HCl and HBr was different between the thermodynamic data and part of the vibrational data, they responded by proposing that the discrepant data features in the vibrational entropy data were less important. This helped Gibbon to dismiss the discrepant vibrational entropy data and progress their analysis to make claims on how the other hydrogen-halides' vibrational entropy contributed to their total entropy. Preserving their frame in this way helped Gibbon to further engage in productive sensemaking, as it helped Gibbon to advance forward in connecting the vibrational entropy values to the total entropy values, which was one of the goals of the analysis.

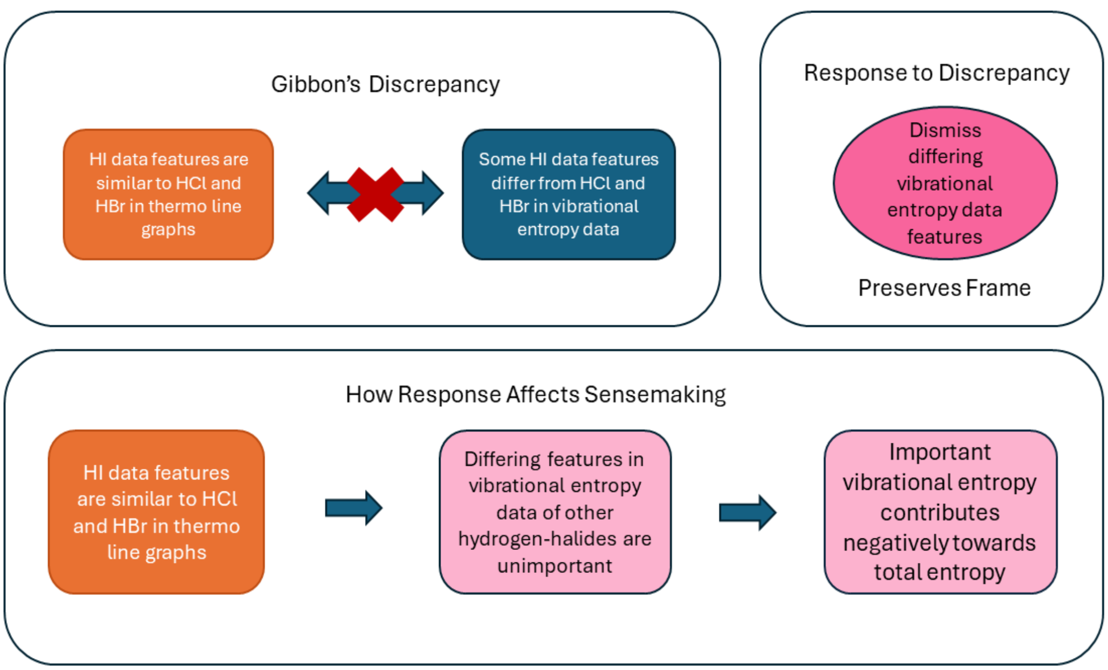

Figure 4. 8: Graphic illustrating Gibbon's discrepancy, response, and sensemaking. <u>Upper left:</u> Gibbon's discrepancy involving a pattern of the other hydrogen-halides in the thermodynamic line graphs (orange) not being replicated in the vibrational entropy data (blue). <u>Upper right:</u> Gibbon preserved their frame (darker pink oval). <u>Bottom half:</u> Gibbon preserved their original reasoning (orange) by reasoning that the discrepant data was less important (lighter pink).

<u>Elaborating the Frame</u>

Many students responded to discrepant data by elaborating their frames. In this, students extended upon their original reasoning; however, students did not dismiss and replace any of their initial observations of the data or any existing parts of their frame. Students who elaborated their frame worked towards incorporating the discrepant data into their explanation to fill the gap in their understanding.

Once students had identified the inconsistency between their frame and the data they observed, students altered their reasoning. Students may have searched and accounted for new data features, such as identifying labels on data features or specific numerical values in the graph. Additionally, students may have incorporated mathematical relationships and mathematical information, usually to identify how different thermodynamic and entropic variables affected each other. Students may also have incorporated new content knowledge or background experimental information into their frame, such as recalling knowledge relating to intermolecular forces or using the visual representations of the crystal structures. Using any of these sensemaking moves helped students to add to their original reasoning and potentially fill gaps in their understanding. For instance, consider Gretna attempting to relate the positive vibrational entropy for HF to its total entropy in the thermodynamic data.

Gretna: *"[HF's] ΔS [vibrational] was positive, this [-TΔS line] should be negative, but it's positive to neutral. Tells me, I guess, the overall the total ΔS [value] would be negative. Okay, okay, we can do this."*

Interviewer: *"Yeah, I can say one thing is that ΔS conformational should be is [sic] negative for every one of them, and it's the same across all the acids."*

Gretna: *"It's just a bigger change compared to the vibrational then, which would make ΔS total negative overall."*

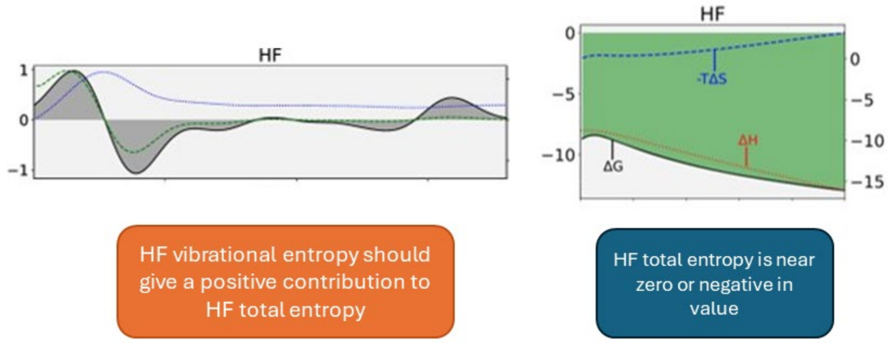

Figure 4. 9: Gretna's discrepancy involving vibrational entropy data and thermodynamic line graph from Bodesheim *et al.* (2020). Gretna's discrepancy involves their prediction that HF's vibrational entropy positively contributing to the total entropy (orange) and their observation that HF's total entropy is negative in value (blue).

Gretna started by predicting that HF's vibrational entropy value should have produced a -TΔS value that was negative to reflect HF's positive vibrational entropy contribution. They also noted that the -TΔS values that they observed for HF were "positive to neutral" which signified a negative value for its total entropy. Gretna responded by taking up new mathematical information from the interviewer, specifically that there was an additional entropy contributor that had a negative value. This allowed Gretna to compare the magnitudes of both entropy contributors and recognize that HF's negative conformational entropy value must have been larger than the value of its positive vibrational entropy so that the sum produced a negative total entropy. Gretna did not need to alter their prior observation of HF's total entropy value, nor did they need to alter their prediction that HF's positive vibrational entropy should contribute positively towards the total. Instead, all they needed to do was incorporate information on the other contributor, so that they could mathematically connect the values of HF's vibrational entropy to that of its total entropy. Gretna's frame elaboration led to more productive sensemaking, as it helped them to accomplish the goals outlined for their analysis. Gretna's frame elaboration helped to establish connections across the entropic data sets and identify a cause for why HF's total entropy was near zero in value.

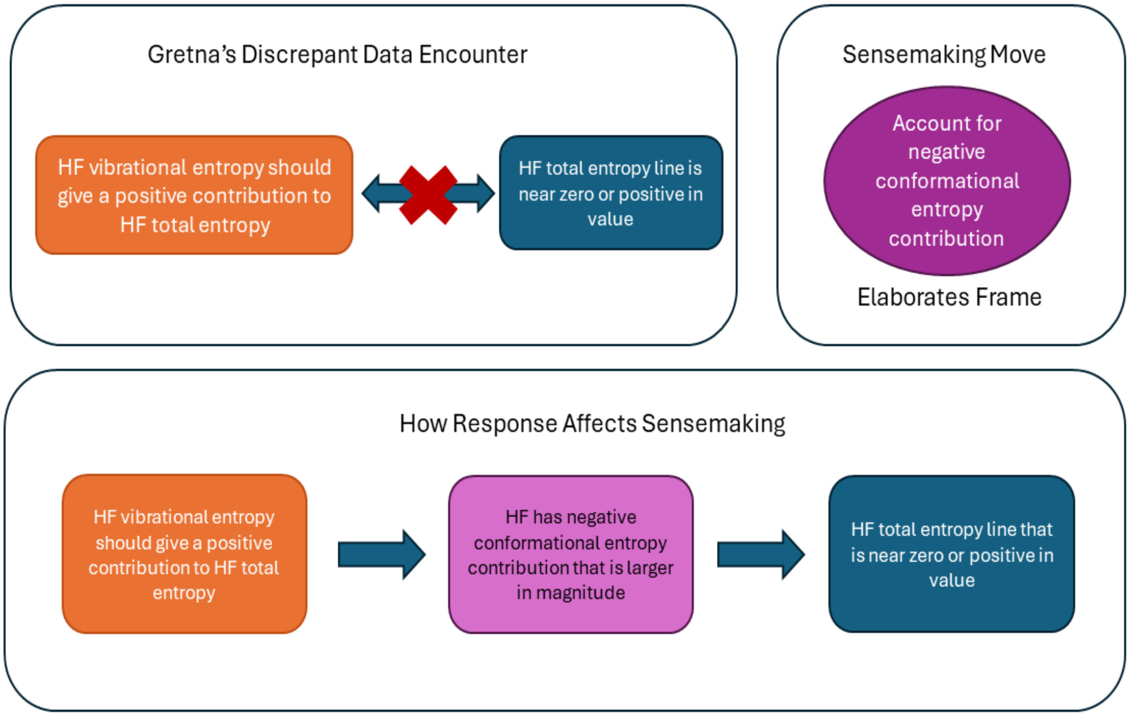

Figure 4. 10: Graphic illustrating Gretna's discrepancy, response, and sensemaking. Upper left: Gretna's discrepancy between their prediction from the vibrational entropy data (orange) and observation of the total entropy (blue). Upper right: Gretna elaborated their frame (darker purple oval). Bottom half: Gretna elaborated on their original reasoning (orange) by incorporating new reasoning (lighter purple), and this helped them to support their other original observation (blue).

Elaborating one's frame did not always contribute toward the most productive sensemaking of discrepant data. Consider Valentine's sensemaking of the intermolecular force data and the vibrational entropy data. When considering the intermolecular force data, Valentine noticed that HF had more non-van der Waals forces. Using this data, they claimed that HF had stronger dipole-dipole forces that caused HF molecules to be held more "tightly" together.

"Because right now, my understanding is sort of like, okay, hydrofluoric acid behaves differently because there's a stronger dipole-dipole force, and potentially also it's just a stronger bond. So there's like, it's like very tight kind of thing close together." (Valentine)

When considering the vibrational entropy data, Valentine also observed that HF 36 had a higher vibrational entropy, which they reasoned meant that in the 36-phase the HF molecules occupied more vibrational states than the 225-phase, and overall exhibited more vibrations.

"If it's here, okay, if it's a positive number that means that the [HF 36] state has a higher [vibrational] entropy, which means that this can occupy more [vibrational] states. So I guess the [36 vibrational] state for the hydrofluoric acid just lets it occupy more regions in space with its vibrations, whereas the 225-state would have fewer regions." (Valentine)

In considering these two data sets and their resulting interpretations apart, Valentine did not seem to encounter any discrepancies. However, when asked how the intermolecular force data and the vibrational entropy data related, Valentine seemed to recognize that their reasoning from each data set was misaligned.

"[HF's vibrational entropy data] It's saying that in the 36-state, there is more possible different, like vibrations that each thing could have. So I don't really feel like I have a

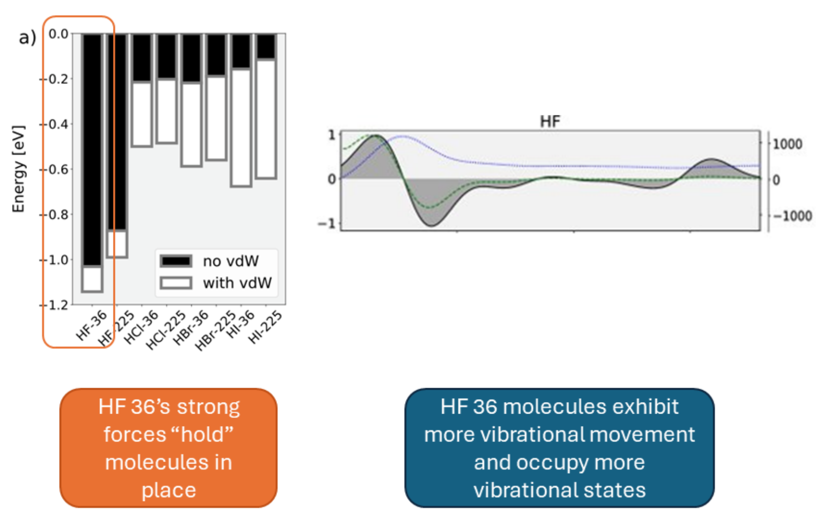

Figure 4. 11: Valentine's discrepancy involving intermolecular force data and vibrational entropy data from Bodesheim *et al.* (2020). Valentine's reasoning of the intermolecular force data that molecules are "held" in place in the 36-structure (orange) conflicts with their reasoning from the vibrational entropy data that molecules are moving (blue).

justification for that because, like the forces are stronger there, so it seems like it should be like stopping it from vibrate – like holding things down in place more." (Valentine)

Here, Valentine seemed to encounter a discrepancy between their interpretations of HF's intermolecular force data and its vibrational entropy data. They stated that HF's stronger forces would predict that the HF molecules should not have vibrated or displayed more vibrations in the 36-state; however, they observed that HF had more vibrational entropy, which meant that HF had more vibrations overall. This discrepancy between interpretations seemed to represent a conflict in their frame. This conflict specifically existed between the frame components "stronger interactions restrict molecules" and "more vibrational entropy means occupying more vibrational states." At this point, it did not make sense to Valentine why something that had strong interactions holding it in place could also vibrate and occupy different vibrational states. In response, Valentine turned to the visual representations of the crystal structures to elaborate their frame and explain why the HF exhibit could have more vibrational entropy in the 36-state rather than the 225-state:

"Okay, here, here's my hypothesis. Oh, overall like because the 225 is like a cube-like – Well, it's more rigid in more different directions, whereas the 36 is like rigid like in this direction ***gestures with hands in x-direction*** *and probably there's like some, you were mentioned like, the up and down thing . . . That's, maybe that's why there's more vibrations because of like that difference."* (Valentine)

Valentine seemed to incorporate the visual representation into their frame which seemed to have activated a new frame component, specifically that "structural rigidity impedes movement." This seemed to guide Valentine's new reasoning, in that the 36-structure's lesser rigidity allowed for more vibrations. This new reasoning allowed Valentine to identify an underlying cause for why they observed more vibrational entropy in HF 36 than in HF 225. Though Valentine's frame elaboration did seem to help to progress their sensemaking to attempt to identify an underlying cause for the vibrational entropy data, the discrepancy remained. Their new reasoning could not address why HF 36 could access more vibrational states when the molecules also had strong intermolecular forces "holding" them in place. By not addressing the source of the discrepancy (the frame conflict), Valentine could not engage in reasoning that could establish connections between the intermolecular force data and the vibrational entropy data. These connections were necessary, as they helped to identify that the stronger intermolecular forces in HF 36 caused its greater vibrational entropy. Thus, Valentine's frame elaboration led to a limited explanation that did attempt to identify an underlying cause, but it did not account for how the different data sets connected together. This meant that Valentine's sensemaking partially achieved the goals for the analysis.

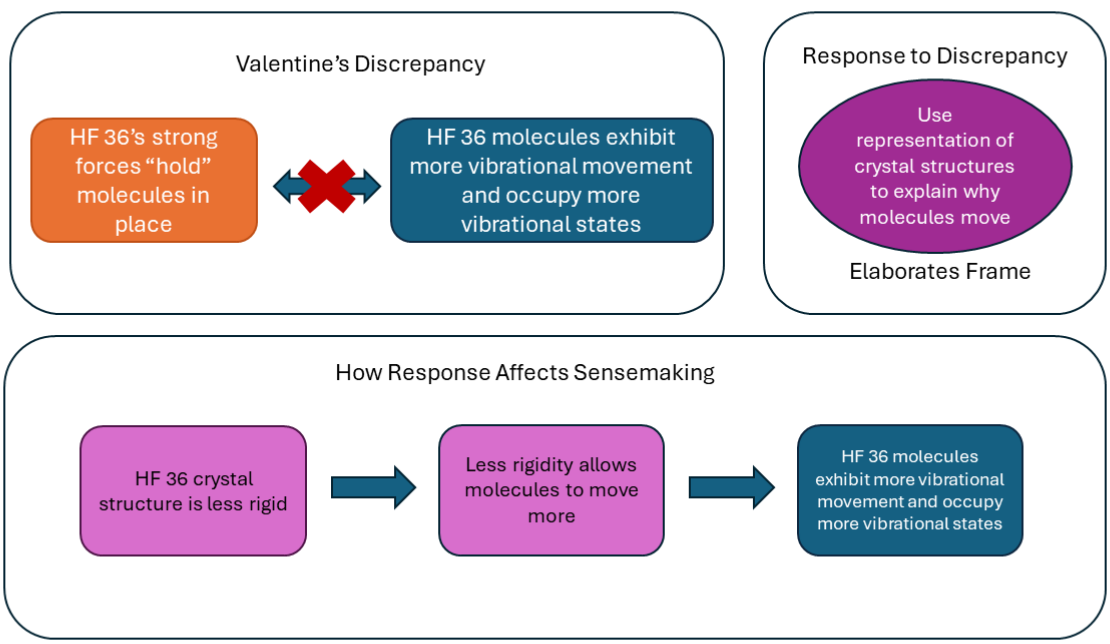

Figure 4. 12: Graphic illustrating Valentine's discrepancy, response, and sensemaking. Upper left: Valentine's discrepancy between their original reasoning of the intermolecular force data (orange) and the vibrational entropy data (blue). Upper right: Valentine elaborated their frame (darker purple oval). Bottom half: Valentine elaborated their frame by incorporating new reasoning (purple) that helped to support some of their original reasoning (blue).

<u>Reframing</u>

Lastly, some students reframed their reasoning after encountering discrepant data. Students who reframed stopped using some component (such as content knowledge or assumptions) or product of their frame (like a prediction) involved in their discrepancy. After this, students worked towards adopting a new perspective to make sense of the discrepancy they had encountered. This new perspective allowed students to start sensemaking with a "blank slate" that could better align with the discrepant data that they encountered.

Once students had identified the inconsistency between the data and their reasoning, they proceeded to question, reconsider, and, in most cases, explicitly dismiss the reasoning involved in the discrepancy. This seemed to indicate that students recognized the limitations of their reasoning, which helped them to recognize why they could not make sense of the data. In doing this, students could then proceed to construct a new explanation that aligned more closely with the discrepant data. Consider Wilber:

During Wilber's explanation construction for the phenomenon, they briefly reviewed the intermolecular force data and pointed out that HF primarily consisted of non-van der Waal forces:

"And what they find is that there is this very large contribution, this very large energetic contribution of non-van der Waals forces." (Wilber)

Wilber also reviewed the vibrational entropy data for HF and explained that its vibrational entropy caused it to favor the 36-phase over the 225, preventing a phase change from occurring.

"But anyway, I think [the vibrational data] seems to be showing that the major difference between the HF versus the other halides is this very large contribution of the vibrational entropy.

And so, it's likely the ability of the molecules in the lattice to vibrate in the orthorhombic phase that is stabilizing it at higher temperatures." (Wilber)

At this point in Wilber's sensemaking, their explanation identified some causes for why HF did not change phase. They identified that HF's intermolecular forces mostly consisted of non-van der Waals forces, but they did not explicitly relate this to any of the other data sets. Additionally, Wilber identified that HF's vibrational entropy was larger in the 36-phase, and they reasoned that this larger vibrational entropy contributed to its stable energy, even at higher temperatures. Because Wilber did not explicitly attempt to establish any connections between these two data sets, they were prompted to consider how the intermolecular data they had reviewed related to the vibrational entropy data. Wilber seemed to immediately encounter a discrepancy at this point. They stated that they would expect something with increased vibrational entropy to be associated with more van der Waals forces but dismissed this prediction immediately.

Interviewer: *"Yeah, how do you think this vibrational entropy is relating to that intermolecular force data that you saw?"*

Wilber: *"Okay, so let's go back and take a look at this [intermolecular force data]. So… it would, hmmm, interesting. Because my first thought would have been that [vibrational entropy] would have increased the van der Waals contribution, but that does not seem to be the case."*

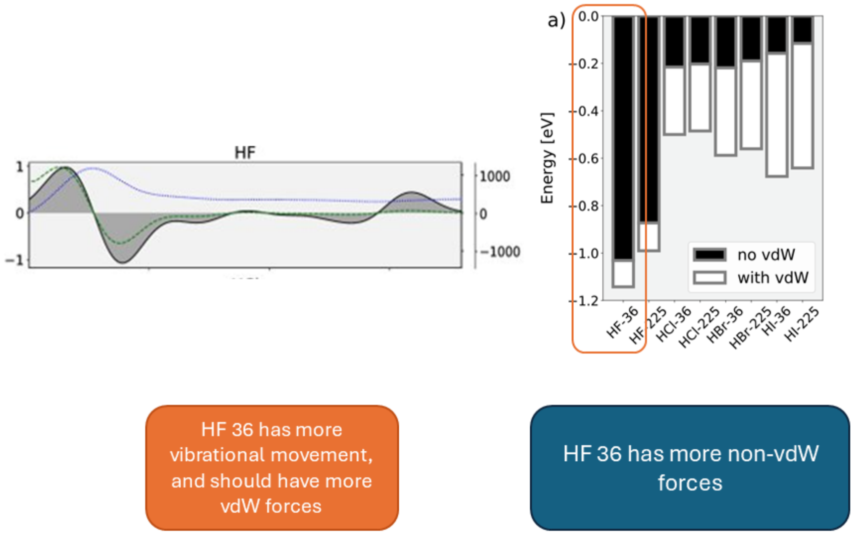

Figure 4. 13: Wilber's discrepancy involving vibrational entropy data and intermolecular force data from Bodesheim *et al.* (2020). Wilber's discrepancy involved their prediction from the vibrational entropy data, that molecules that moved more should have more van der Waals forces (orange), not being what they observed in the intermolecular force data (blue).

Wilber's discrepancy between their prediction and the data was similar in nature to Valentine's (and many other students') discrepancy. Their frame seemed to anticipate that a substance's strong intermolecular interactions would impede movement and vibration or conversely, that a substance capable of vibrational movement would have weaker intermolecular forces holding the molecules together in a solid. As soon as Wilber voiced their prediction, they immediately stated that it did not align with what they observed in the data. This could demonstrate Wilber dismissing the component of their frame that shaped their prediction. By dismissing this expectation of what should be in the data, they essentially could start attempting to construct a new explanation that was more closely aligned with the different data sets.

"*And so... you know, it could just be if these [molecules] are vibrating, they're moving around in space a little bit. I guess these would be vibrating, and so [the vibrations] would change kind of this hydrogen bond interaction as well. And it would change, you know, the* *dipole forces that would be occurring, so that would be changing the magnitude of those [dipole forces].”* (Wilber)

After dismissing their prediction and the parts of their frame that shaped that prediction, Wilber adopted a new perspective and attempted to make sense of how the vibrational entropy HF experienced in the 36-phase affected its intermolecular forces. They specifically reasoned that the vibrational movement affected the magnitude of intermolecular forces. Then Wilber went on to consider how the intermolecular forces might have conversely affected the vibrational entropy.

“*I mean, it could also be that there, these hydrogen, it could kind of be the other direction as well. In that, the hydrogen bonds and the dipole interactions are kind of maybe stretching these bond lengths to some extent or causing them to be in a, in sort of an unusual state, maybe stretched or compressed, or whatever. And that could be increasing the amount of vibrations that are required, or that occur as a result of that. And so, it could actually be that the vibrational entropy is sort of a result of these dipole interactions or the [hydrogen] bonds*.” (Wilber)

Here, Wilber established connections between the intermolecular force and vibrational entropy data sets and integrated their findings into a causal explanation. They reasoned that the intermolecular interactions might have caused the vibrational entropy they observed, specifically pointing out that the intermolecular interactions between molecules might have also caused the intramolecular bonds to stretch. In turn, this would have caused the atoms to move, which would be associated with the vibrational entropy data. In adopting this new perspective, Wilber could effectively establish connections between the vibrational entropy and intermolecular force data sets and identify an underlying cause for HF's vibrational entropy. Thus, Wilber's reframing contributed towards productive sensemaking that fulfilled the goals of the analysis.

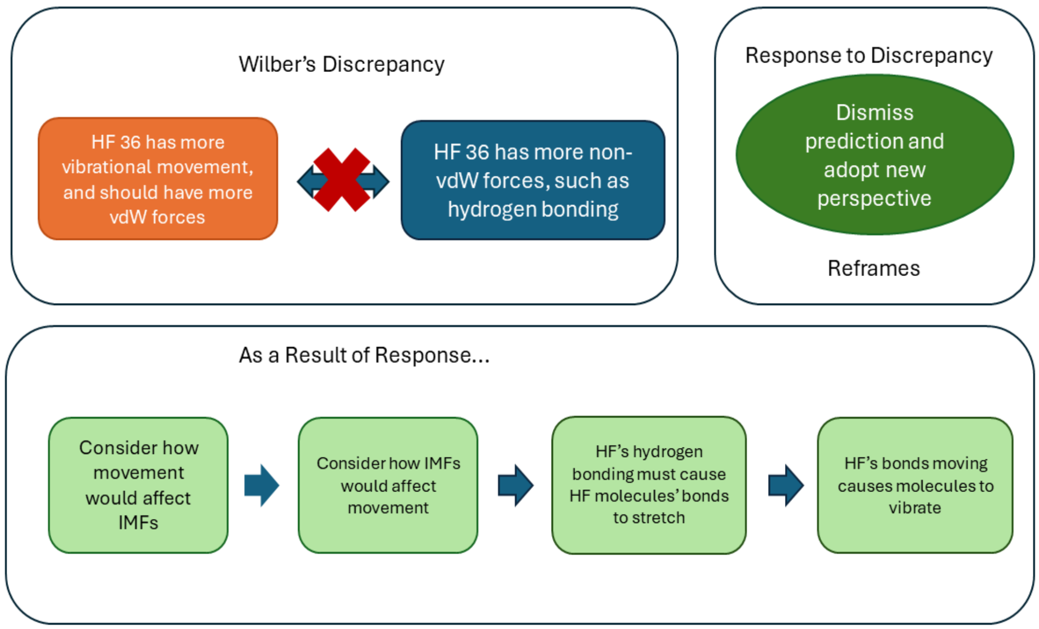

Figure 4. 14: Graphic illustrating Wilber's discrepancy, response, and sensemaking. Upper left: Wilber's discrepancy between their prediction that HF 36 should have more non-vdW forces (orange) and their observation of the intermolecular force data (blue). Upper right: Wilber reframed (darker green oval). Bottom half: Wilber reframed and adopted entirely new reasoning (lighter green).

Discussion

The current study explores how different chemistry graduate students respond to and make sense of data that they do not expect to observe. The analysis identified three distinct responses that students gave while analyzing multiple data sets to construct an explanation for a chemical phenomenon. These responses differed to the extent that students altered their reasoning and incorporated the discrepant data into their explanations. Students who did not alter their reasoning preserved their frame. Students who elaborated their frame did alter their reasoning by expanding upon it in some way. Lastly, students who reframed their reasoning dismissed some parts of their reasoning and adopted a new perspective to account for the discrepant data.

This work provides evidence that there is not necessarily one "correct" response to discrepant data. The same response could lead to explanations of varying productivity, such as what can be observed when considering both Fremont and Gibbon's frame preservations.

When Fremont preserved their frame, it seemed to stop from engaging further with the intermolecular force data set. By dismissing the HF 225 intermolecular force data, Fremont limited what data they could use to establish a causal explanation. This led to Fremont constructing a partial causal explanation that only considered the vibrational entropy and thermodynamic data. In this way, Fremont's response to preserve their frame seemed to contribute towards less productive sensemaking, as they did not attempt to connect any of the intermolecular force data to other data sets, nor did they identify any underlying causes for what they had observed in the other data sets.

Conversely, when Gibbon preserved their frame, it seemed to help them to move on to establish how the thermodynamic and vibrational entropy data sets related. In preserving their frame, Gibbon spent less time scrutinizing a data feature that did not seem relevant to advancing the goals of the analysis. Gibbon could then return to analyzing other data features that helped them to establish connections between the vibrational entropy and thermodynamic data.

Both Fremont and Gibbon responded to discrepant data features by preserving their frame, but they used their frame preservation in differing ways. Fremont stopped engaging with the whole intermolecular force data set and stopped their sensemaking entirely, whereas Gibbon stopped engaging with the specific data feature and continued establishing connections between different data sets. Given this, one's response to discrepant data is best used when it helps to align one's sensemaking activities (i.e., what one does with their reasoning) with the specific goals of one's analysis.

Additionally, this study showed that students who encountered the same or similar discrepancy sometimes often used different responses that progressed their sensemaking in different ways. Both Valentine and Wilber experienced similar discrepancies when asked how the intermolecular force data related to the vibrational entropy data. Both students experienced some sort of discrepancy between their interpretations of the two data sets, which likely resulted from a conflict with the components of their frames. Valentine responded to their discrepancy by expanding upon their reasoning by using the visual representations of the structures to identify a cause for why the HF 36 had more vibrational entropy. Although this elaboration did help them to identify a cause for HF 36's larger vibrational entropy, their explanation could not account for how the intermolecular force data affected or related to the vibrational entropy data. This could be because the elaboration did not address the underlying conflict within their frame, it could only "add on" to existing reasoning. In contrast, Wilber's response to the discrepancy was to reframe their reasoning by dismissing their prediction and constructing an explanation that could relate the two data sets together. By dismissing their expectation for what they believe the data should have been, Wilber could essentially construct a new frame that could fully account for both the vibrational entropy and intermolecular force data. This response to the discrepancy helped progress Wilber towards more productive sensemaking, as it actually helped to establish a relationship between the two data sets.

Conclusions

From this study, graduate students had three responses to discrepant data they encountered while analyzing multiple data sets to explain a chemical phenomenon: preserving the frame, elaborating the frame, and reframing. For descriptions of each response, see Table 4.1. When students preserved their frame in response to the discrepant data, they made no changes to their reasoning. Some students went so far as to dismiss the discrepant data feature involved in their discrepancy. Students who elaborated their frame in response to the discrepancy did change their reasoning, but the changes came from incorporating new conceptual knowledge or background experimental information, searching for more data, or using mathematical information. Elaborating their frame in this way expanded students' reasoning, but it did not replace any existing components of their frame or any initial observations. Finally, students who reframed in response to their discrepancy dismissed some aspects of their frame, which allowed them to start constructing an explanation that more closely aligned with the discrepant data.

All three responses had varied effects on students' sensemaking. This meant that there was no one "correct" response to discrepant data. Instead, responses that were more productive for sensemaking were responses used to fulfill the goals of the analysis. For this task specifically, responses were most productive when they helped students to identify underlying causes for the chemical phenomenon or establish connections between the data sets. All three responses to discrepant data were capable of progressing students' sensemaking to accomplish these goals; however, some responses were more appropriate depending on what was involved in the discrepancy (e.g., what data features, frame components, etc.).

Table 4.1: Description of different graduate student responses to discrepant data

Type of Response to Discrepancy	**Description of Response**
Preserving the Frame	After identifying the discrepancy, students preserve their reasoning in some way. To do this, students might stay "stuck" or simply move on to analyzing other data features. Some students explicitly dismissed the discrepant data feature.
Elaborating the Frame	Once students identify the discrepancy, students incorporate new elements into their reasoning. Students may search for more data features, use mathematical information, or incorporate new content knowledge or background experimental information. In this way, students' reasoning could be expanded upon, but their original frame components remained intact. Students also did not replace any previous observations made of the data.

Reframing	In response to the discrepancy, students dismiss a component of their frame. This gives students a “blank slate” to build a new explanation that more closely aligns with the discrepant data feature.

Implications for Practice

The findings indicate that graduate students respond to unexpected data in a variety of ways when analyzing multiple data sets to construct an explanation for a chemical phenomenon. Additionally, all three responses were capable of progressing students toward more productive sensemaking, but they did not always do so. Given this, it stands to reason that chemistry graduate students could use guidance in learning when and how to use the different responses to unexpected data. Graduate advisors play a critical role in their graduate students’ success (Mason, 2012). Researchers have also suggested that advisors can help their graduate students develop autonomy by establishing explicit expectations for students (Barnard & Shultz, 2020). One way in which advisors can do this is by helping their graduate students learn the different ways to respond to discrepant data and identifying when different responses may be most appropriate to use.

To start, advisors can help students identify the purpose of their data analysis and identify what goals must be achieved to fulfill the purpose. Goals could include identifying underlying causes for a phenomenon, using data to determine optimal experimental conditions, etc. In setting goals for their analyses, students establish guidelines to use when making decisions in their analysis.

Additionally, advisors can outline a mental schema to help guide students through a series of steps when deciding how to respond to discrepant data. This schema could take the form of a conversation with the student, or it could even take the form of a written handout for students to keep while analyzing data.

As a first step, students can consider if engaging further with the discrepant data is necessary to fulfill the purpose of their sensemaking. In other words, students should decide if time could be better spent moving on to analyze other data features or other data sets. In some instances, it could be more productive for students to preserve their reasoning if the discrepant data is not needed to meet the criteria of their analysis.

If students decide that the discrepant data is important to fulfill the goals of their data analysis, students can consider if they are missing any details in their sensemaking. They can attempt to elaborate their frame by searching for and incorporating other data features that they may have missed, incorporating any relevant content knowledge that may assist in making sense of the data, or considering information such as mathematical equations or other experimental information. Elaborating their frame in this way can fill “gaps” in students’ reasoning, which may lead to more productive sensemaking when students’ initial reasoning might have missed a detail in the data or background information.

In cases where students have attempted to elaborate their frame and still have not fulfilled their goals for the analysis, students can consider reframing their reasoning. Here, students can explicitly identify what aspect of their reasoning is discrepant with the data, and search for alternative frames that could be used to make sense of the data. This would entail adopting an alternative perspective that could be used to more productively account for and make sense of the discrepant data. This step is far more exploratory in nature and could require more time and cognitive energy to fully make sense of the discrepant data. Thus, reframing should be considered as a response to discrepant data only when students have deemed the discrepant data necessary to make sense of and they have exhausted their ability to build off their original reasoning.

Limitations

The findings from this study are specific to the design of the task, the research methods used, and the analytic choices made.

First, the task utilized different visual representations of data, mainly in the form of graphs, from a recent publication in a physical chemistry journal (Bodesheim et al., 2020). It should be noted that the publication features “clean” published data that is not representative of most research contexts where students collect raw data that is far messier in nature. Analysts are likely to encounter a plethora of different discrepant data when collecting and making sense of raw data, and they may require different responses to fully make sense of the data. Some studies within physics education and science education have investigated this within the context of undergraduate laboratories (Adams, 2023; Crujeiras-Pérez & Jiménez-Aleixandre, 2019; May et al., 2020, 2022). However, these studies are contextualized in introductory undergraduate courses that use scaffolded experimental procedures for students to follow. Graduate students must often modify or construct their own experimental procedures for their research, which is likely not captured in the previous studies. I suggest future studies explore how the sensemaking of experimental data is influenced by how graduate students plan and carry out investigations of chemical phenomena.

The study’s methods also employed semi-structured interviews in which graduate students primarily worked alone. It is more likely than not that graduate student researchers analyze and make sense of data in accompaniment with others, such as fellow student researchers, postdoctoral researchers, and research advisors. The teaching interview format captured some ways in which students, such as Gretna, took up information provided by the interviewer to alter their reasoning in some way. Additionally, previous work identified that some undergraduate students altered their reasoning of data when exposed to other perspectives, such as through peer review (Berg & Moon, 2023). Although the findings are specific to undergraduates, it is still possible that making sense of data in accompaniment with others would expose graduate student researchers to alternative perspectives that they might consider in their analysis. Thus, future research should consider this additional social dimension when graduate students make sense of data with others.

Finally, one of the analytic decisions was to carefully select episodes in which graduate students voiced their discrepancies aloud during their data analysis. There are likely moments in which participants did not voice their discrepancy and engaged with it internally, which cannot be captured in the think-aloud interview format. There were also moments in which students did not seem to recognize that their reasoning was not aligning with the data. I chose to omit these moments from our analysis, as the goal was to capture how students respond when they notice that their reasoning is discrepant with data. This necessitated some step of identifying an inconsistency that omitted episodes did not contain.

CHAPTER 5

CONCLUSIONS

Chemistry and science educators have long called for undergraduate STEM courses to incorporate more opportunities for students to engage in science practices (Cooper et al., 2015; National Research Council, 2012; Sharon & Baram-Tsabari, 2020; Talanquer & Pollard, 2010). At the center of all eight science practices is empirical evidence, which means students need to develop competency in engaging with and making sense of empirical data. However, there are several challenges that students experience with handling data (Becker et al., 2017; Bolger et al., 2021; Heisterkamp & Talanquer, 2015; Kanari & Millar, 2004; May et al., 2022; Meister et al., 2021; Phillips et al., 2021; Zagallo et al., 2016), and there are limited classroom activities designed to help students develop their competencies (Bolger et al., 2021; Zagallo et al., 2016). To that end, this dissertation seeks to characterize how undergraduate and graduate chemistry students engage in data analysis and investigate how peer review can be used as a classroom activity to help undergraduate students develop their data analysis competency. I conducted a total of three qualitative studies in which I interviewed undergraduate and graduate chemistry students as they engaged in different data analysis activities, with the main findings highlighted below:

1. **Prior knowledge and experiences help to shape undergraduate students' frames when making sense of data.**

Graph comprehension and data analysis are both vastly influenced by one's prior knowledge and experiences (Becker et al., 2017; Carpenter & Shah, 1998; Heisterkamp & Talanquer, 2015; Pinker & Feedle, 1990; Shah & Carpenter, 1995; Shah & Hoeffner, 2002). Given that students must engage with empirical data, often in the form of graphs, there is a need to understand how prior knowledge and experiences interact with their engagement in science practices with graphical data. My study in Chapter 2 characterized how undergraduate students activate and use their prior knowledge as they make sense of data.

I first interviewed undergraduate students as they analyzed a graph to determine optimal experimental conditions and then compared their responses to alternative viewpoints through a simulated peer review. I then deconstructed students' analyses using Data-Frame Theory to identify what prior knowledge and experiences students used and characterized how it was used throughout the students' analysis. My findings suggest that students first engage in making sense of data by establishing a frame, which may be composed of prior knowledge or experiences. This frame then helps students to filter what data they notice and interact with. There are several ways in which the data and frame interact, such as when the frame is used to evaluate the data or when the data causes a shift in frame. Some students' frames also changed when students considered alternative perspectives in a simulated peer review.

My work suggests how data analysis tasks are written and scaffolded can prompt students to activate prior knowledge and experiences that can be productive in framing students' data analysis. Additionally, students' frames may change when considering others' perspectives through activities such as peer review.

2. **Social comparisons against sample responses can serve as a vehicle to prompt undergraduate students' self-evaluation and help students generate internal feedback about their performance.**

Data analysis is a difficult competency to develop within the undergraduate chemistry classroom, and there are a limited number of documented classroom interventions that have been designed to advance students' competencies (Bolger et al., 2021; Zagallo et al., 2016). To address this, the study in Chapter 3 investigated how using a simulated peer review could prompt undergraduate students to self-evaluate their data analyses and generate internal feedback for themselves.

I interviewed general chemistry students as they compared and evaluated their data analyses from a task against pre-constructed sample responses. Using these interviews, I applied social comparison theory and internal feedback theory to generate a model to document how students evaluated their analyses and explain why they may have chosen to revise their work. The model shows that the comparison against a sample response could prompt students to self-evaluate their work and generate internal feedback. Depending on the presence and nature of the internal feedback, students might choose to revise or maintain their work. Critical internal feedback informed students that their work did not fully meet the criteria for success for the task, and the feedback motivated students to make revisions of some kind in response. Whereas a lack of critical internal feedback validated students' work and informed students that they met the criteria for success, and they did not feel the need to revise.

My work suggests that simulated peer review or comparisons more generally (such as against a rubric or exemplar) can be used to help undergraduate students self-evaluate and generate internal feedback for their work. Instructors can use preconstructed responses to communicate criteria or more nuanced ideas for students to tease out and incorporate into their work. This could be especially useful for large-enrollment classrooms in which instructors may not be able to give individual feedback on students' work.

3. **Graduate students respond to discrepant data in a variety of ways, some responses may be more appropriate than others depending on the context in which they are used.**

As part of their training and education, graduate students must develop to become competent and independent researchers who contribute new knowledge to their field (National Academies of Sciences, 2018; The American Chemical Society, 2012). As a part of advancing the knowledge of their field, students must engage with uncertainty and learn when and how to best engage with data that is discrepant with their expectations (Grolemund & Wickham, 2014). Though this is important, it is not known how graduate students respond to data that is discrepant with their expectations.

Thus, I interviewed chemistry graduate students as they analyzed multiple data sets to explain a chemical phenomenon. I identified the various moments in which students encountered data that they were not expecting to see (i.e., discrepant data). I characterized both the components involved in students' encounters with discrepant data (e.g., what data was involved, what knowledge was involved, etc.), characterized how the students made sense of the discrepant data, and evaluated how students' responses affected their sensemaking that followed. The results of my analysis in Chapter 4 suggest that chemistry graduate students respond to discrepant data in three different ways. Additionally, all three responses could contribute towards productive sensemaking for the student, if they were used to help students meet the goals of the analysis. However, my findings also suggest that graduate students may need additional guidance in learning how to respond to discrepant data to best meet their goals.

Future Directions for Research

My work on generating internal feedback is highly contextualized in the students' social comparisons against sample responses in a simulated peer review. The processes characterized could also be present in contexts in which students compare their work to rubrics or exemplars. Nicol recently identified similar processes of internal feedback when undergraduate students compared their written work to exemplars (Nicol & Kushwah, 2023); however, more research is needed to investigate if internal feedback is generated when comparing to rubrics. Additionally, other contexts likely generate internal feedback, such as receiving feedback from others. This is an especially important context to consider, as students will encounter this form of feedback in academic environments and the world beyond. Future research should explore how internal feedback is generated when students receive feedback from others.

Next, the results of my research are primarily grounded in how students interact with chemical data in the form of graphs. I chose to use graphs because chemistry students are more than likely to have interacted with graphs before, and students possess very rich graphical schemas that help them navigate their graph interpretations; however, there are many other forms of data that chemistry students will interact with, such as data collected from spectroscopy or chromatography. These forms of data require specialized knowledge; thus, students may interact with the spectra and chromatographs in different ways than graphs. There is some research on how students make sense of certain spectroscopies (Cartrette & Bodner, 2010; Connor et al., 2021; Connor & Shultz, 2021), but many of these studies tend to disconnect the use of spectroscopy from data analysis as a science practice. That is, the studies focus on how students use their conceptual knowledge to read spectroscopy, but they do not explicitly consider how students use the data to generate claims or construct explanations, which is a necessary aspect of practicing science. Thus, future research is needed to explore how students engage in data analysis with other common forms of chemical data.

Additionally, the tasks that I developed have only used published chemical data to investigate how students generate claims or explanations from the data. In using published data for the tasks and focusing on claim generation, I could not investigate how collecting and using raw experimental data affected students' analyses. It is more likely than not that collecting and engaging with raw experimental data will elicit different sensemaking processes that my studies could not capture. Several studies in science education, physics education, and biology education have implemented this facet of data analysis into their studies (Adams, 2023; Bolger et al., 2021; Crujeiras-Pérez & Jiménez-Aleixandre, 2019; May et al., 2022), but there remains a need to also investigate this within the context of chemistry. The studies outside of chemistry cannot account for students analyzing discipline-specific sources of experimental data, such as chromatography and spectroscopy, which require some sort of conceptual knowledge to make sense of.

Finally, the findings from Chapters 2 and 3 suggest that some students may alter their reasoning when exposed to others' perspectives. It is possible that this could also occur in social contexts other than peer review, such as working with peers in a chemistry laboratory course or working with a mentor. In working with others, students may need to consider different priorities in the analysis (e.g., considering others' priorities when making a decision from evidence) or consider others' prior knowledge and experiences with chemical data. How students consider these alternative viewpoints is very likely influenced by different social and interpersonal factors, such as power dynamics between those involved in analyzing the data, how students respond to others' contributions to analysis, etc. Future research should investigate how this additional social dimension shapes how students may make sense of data with others, whether in groups of peers or with mentors.

APPENDIX A: INTERVIEW PROTOCOL FOR CHAPTERS 2 AND 3 AND SAMPLE RESPONSES

We are going to do this interview in two parts. In the first part, you will look at a graph with a question and make some claims about it. In the second part, you will evaluate other students' arguments about that graph. In the second part of the task, the student will evaluate <u>three</u> sample arguments.

There is no right or wrong answer here, we want to understand how you are thinking. So, though it may feel awkward, try to do all your thinking out loud so we can understand. We will also prompt you for more thinking throughout.

<u>Part one:</u>

Screenshare document with student. Scroll up for student to read the task. At any point that the student pauses for long periods of time or seems confused, remind them that they can reread the task.

Check to make sure that they can see the graph clearly.

I. Graphing warmup
    a. Look at the x-axis. What is shown on the x-axis?
        i. What's changing?
    b. Look at the y-axis. What is shown on the y-axis?
        i. What's changing?

II. Chemistry context warmup
    a. What are you noticing with the graph? Annotate the screen to mark the parts that you noticed
        i. Ask about slopes, peaks, points, etc.

III. Between 0 M, 2 M, and 4 M, what concentration would you recommend using?

IV. What information helped you reach your answer?

  a. What kinds of data did you use from the graph? (Slope, points, ratios, peaks, line trends, etc.)
  b. What specific data did you need to use from the graph?
  c. What details from the task did you need to help you make a decision?
  d. What information on the graph did you not use?

V. How does the information you used support your conclusion?

  a. Why was this information important for you to use?
  b. What choices did you have to make to come to your conclusion?
  c. How would the waste affect the extraction?
  d. How would your choice of concentration pan out in an extraction compared to another concentration?

VI. Write the argument/response in the chat and summarize what you've said out loud. Include the details that you think are relevant to convince the scientists behind the project.

Part two: Scroll down and show the empty graph and prompt on the screen. Remind students that they don't have to make a graph right away and that the experiment they propose is completely up to them.

VII. How would you design an experiment to consider the role time plays?

  a. What variables would you need to consider?
  b. What would you assess with this experiment?

VIII. What information would you try to get from this test that would support your choice of HCl concentration? (show the empty graph and test on screen)

   a. What are some of the advantages of the test?

   b. What information could this test give to support your answer?

   c. How would that information support your answer?

IX. Draw an outcome of the experimental information you would obtain if you performed the test. Be sure to label the axes

   a. What's happening in the graph?

   b. Walk me through what you're showing in the graph

<u>Part three:</u> Copy and paste the responses one at a time in the chat. Ask them to read it and give a thumbs up when they're done reading it.

I. What are some things that you're noticing with this response?

   a. What information did they pull from the graph?

   b. What information did they leave out of their response?

II. How convincing do you find this response?

   a. What are some strengths of the response?

   b. What are some weaknesses of the response?

   c. What changes would you suggest for the response?

III. How does this response compare to yours?

   a. In what ways does their response affect your own thinking?

   b. What is your confidence in your own response now? Why?

IV. Out of the three responses we've shown you, which one do you find the strongest? Why?

Below are the three sample responses that students evaluated during the peer review stage:

4 M sample response

Starting at 0 M HCl, Au is only 65% extracted. As the concentration of HCl increases, so does the percent extraction of Au. The percent extraction of Au at 4 M HCl has reached its maximum and is 95% extracted. Because there is a large percent of Au in the organic layer at 4 M, the best concentration of HCl to use is 4 M.

0 M sample response

The best concentration of HCl to use is 0 M HCl. The percent of waste extracted increases as the concentration of HCl is increased. Because the waste extracted is not minimal at 2 M or 4 M, 0 M is a better option. 0 M HCl also has 65% Au extracted, so there is a large amount of Au that can be gathered with no waste present.

2 M sample response

From 0 M to 4 M HCl, both Au and waste extraction increase as the concentration of HCl increases. Therefore, the best concentration of HCl to use is 2 M HCl. The increase in Au extraction from 2 M to 4 M does not justify the increase in waste. But the increase in Au extraction from 0 M to 2 M does justify the increase in waste.

APPENDIX B: CHAPTER 4 INTERVIEW PROTOCOL AND WALKTHROUGH OF GRADUATE STUDENT TASK

I. Slide I

For this study, you've consented to participate in a think-aloud interview. This means that you will need to say what's on your mind when you're reasoning through this task. This could feel a little awkward, but I will be asking questions to help you do this. The key to this is that if a thought or idea pops up, I want you to verbalize it. If any of my questions don't make sense or you need to hear them again, please let me know.

The goal of this study is to see how you make sense of and use data to construct an explanation. You're going to be seeing four different datasets today from a recent ACS publication, and we will be going through all of them. You are not being assessed on your explanation at all. What I am trying to see is how you go through the data and reason with it, not whether you produce the same explanation as the authors. We can also go over the author's explanation, but during the interview I want you to do your best to use your knowledge and reasoning.

During the interview, you are going to be in control of the screen and slides. That means that you get to go back and forth on slides if you want, and you can draw on the screen to take notes or mark certain things. In fact, we're going to strongly encourage you to mark and annotate the screen as you go, so I can have a reference for what you're seeing when I analyze the interviews.

What questions do you have for me before we start?

Press record

II. Let's practice thinking out loud with a question really quick: What do you research? What does your research focus on?

In past interviews, many students have found it helpful to be taking notes as they go through the task. We recognize this is a bit harder to do via Zoom, so we want you to take notes that you need on some separate software or app that you can share with us later. We would like you to take notes and write down things as needed. If you have a tablet and want to draw on the screen, that is totally fine as well.

III. Slide II

So today we're going to be talking about a phase change in four different binary acids. These binary acids are HF, HCl, HBr, and HF. At lower temperatures, these acids tend to take an orthorhombic crystal structure. The orthorhombic crystal structure is Figure A on the left. Here the small atom represents the hydrogen, and the larger atom represents the halogen. The black and white in this figure represent the three-dimensional orientation. So the black rows stack on top of each other, and the white row sits behind the black rows. What questions do you have about this structure?

As temperatures increase, the binary acids can undergo a phase change. You might be more familiar with phase changes like solid to liquid, but we can also have phase changes where the crystal structure of a solid changes. The new crystal structure that the acids form is a cubic crystal structure, which we can see in Figure B. The black atoms represent the halogen and the white atoms represent the hydrogens. What questions do you have about the phase change or cubic structure?

IV. In this paper that we're using data from, a group of physical chemists experimentally documented the orthorhombic and cubic crystal structures of the four binary acids across a specific temperature range. We still have the two crystal structures at the top of the screen, and then below that we have a bar graph of the different structures. Just above the HF bar, there is a legend to help with the color coding.

What are some of the things that you're noticing in this bar chart?

a. What crystal structures are present for each of the acids?

b. What phase changes are happening for the acids?

c. What are some interesting things you can see in this bar chart?

d. What do you think warrants investigating further?

e. Optional: What are some "big-picture" or "bird's eye view" conclusions you can make from seeing this bar chart?

V. To investigate this further, the physical chemists decided to simulate the acids in two structures. So from now on, when you see the 36, it refers to the orthorhombic structure. And if you see the 225, it refers to the cubic structure. While controlling for the acids' structures, the chemists performed a few different computations. In this first one, we're going to see the energy associated with each acid for the different acids.

They characterized the kind of intermolecular force in two ways: van der Waals and non-van der Waals. So the white "with vdW" means van der Waals forces' contributions. And the black "non vdW" represents other types of intermolecular forces' contributions.

What are some of the things that you're noticing in this bar chart?

a. What is on the x-axis?

b. What is on the y-axis?

c. What other types of intermolecular forces do you think are present?

d. How are the energies of the acids comparing across the chart?

e. What are the contributions of the vdW forces for each acid?

f. What are the contributions of the hydrogen bonding for each acid?

g. What do you find interesting about what you're seeing in the bar chart?

VI. Now we're going to cover a simulated phase change from the 225 cubic to the 36 orthorhombic. The physical chemists simulated this phase change across a range of temperatures and found the thermodynamic values associated with the change. What you're seeing plotted on the graph is the change in a thermodynamic value because we're dealing with a phase change. And remember that in thermodynamics, we denote a change mathematically as the final state minus the initial state.

What are some of the things that you're noticing with the line graphs?

a. What colors are you seeing for each graph? What do they mean?

b. What are you noticing with the black lines for each acid? What does it represent?

c. What are you seeing with the blue lines for each acid? What do they represent?

d. What are you noticing with the red lines for each acid? What do they represent?

e. How are the slopes comparing for the ΔG lines?

f. How are the slopes comparing for the ΔH lines?

g. How are the slopes comparing for the $-T\Delta S$ lines?

h. How do the different lines relate to each other?

i. What information do you think the line graphs are telling you about the phase change from 225 to 36?

j. What are some interesting or unique things that you're finding from the line graph?

VII. Something that the physical chemists wanted to investigate further was the entropy associated with the phase change. How we commonly think of entropy is through the idea of "disorder." This type of entropy is known as conformational entropy, and it is one kind of entropy that we know contributes to the phase change. Because all of the acids are undergoing the same phase change from 225 to 36, they are experiencing the same change in conformational entropy. We also have another form of entropy that contributes to the total entropy change called vibrational entropy. We'll go over what it entails in the next slide, but all you have to know now is that it's another source of entropy. How are you thinking about the total entropy change now that you know that there are two sources of entropy?

a. Why do you think that the different acids' ΔS slopes are different from each other?

b. What do you think is making them different?

VIII. Vibrational entropy is probably a new concept, so we'll cover the basics of it before we start considering the data here.

Vibrational entropy is a form of entropy that can be encountered in a lot of solid-state materials. As we know, atoms are always vibrating because they have thermal energy. The idea behind vibrational entropy is that the atoms in the solid structure will vibrate and through these vibrations, they will explore the space around them. The more space that is being explored, the stronger the vibrational entropy. So then when we're measuring a change, we're looking at the difference in the space explored via vibrations.

To solve for the vibrational ΔS, we have to do a series of calculations for the phonon density of states or pDoS. That's what the black line represents. We are going to ignore the black and green lines on this chart, and just focus on the blue line for this chart. The phase change was simulated at 300 K, and then the physical chemists looked for the strength of the vibrational entropy contributions across different frequencies of vibrations.

What are you noticing in this chart?

a. What is happening for each of the acids?
b. What are some interesting/unique things happening in the vibrational entropy chart?
c. What do you think the strength of the vibrational entropy is for each acid?
d. How are you thinking about the change in total entropy now?
e. How do you think the vibrational entropy change contributes to the total entropy change for each acid?
f. How do you think the vibrational entropy contributes to the change in Gibbs Free Energy?

IX. Now that we've gone through the four datasets to explore the crystal structures for the acids, I want you to explain why the HF does not take a cubic structure in this temperature range like the other acids do. You can use whatever data you think is necessary and you are encouraged to revisit some of the old slides to refresh. Why do you think that the HF only stays in the orthorhombic phase?

a. If they only bring up thermodynamic data, why do you think these thermodynamic values are supporting that orthorhombic state? What is it making it favor that state?

b. If they only bring up IMFs, how does that relate to the thermodynamic data you saw?

c. How does the data support your explanation?

Once the student is finished, end the recording

<u>Description of Task</u>

The first data set included a calculation of the energy involved in the different non-covalent interactions (i.e., intermolecular forces) of the two structures of the hydrogen-halides. The data depicted the energies of both the van der Waal forces and non-van der Waal forces of each structure of the hydrogen-halides. From this data set, the author concluded that HF's intermolecular interactions were dominated by hydrogen bonding, with significantly more hydrogen bonding occurring in the 36-phase. In contrast, the other hydrogen-halides' intermolecular forces were much weaker in comparison and were predominantly made up of van der Waal forces.

In the next data set, the authors' computations compare the thermodynamic states of each structure for the different hydrogen-halides. This involved calculating the difference in three thermodynamic states across different temperatures: the Gibbs free energy, the enthalpy, and the entropy. In this data set, HF consistently shows lower Gibbs free energy and enthalpy in the 36-structure, and relatively no difference in the entropy for the two states. Although the other hydrogen-halides do favor the 36-structure at lower temperatures, they all reach points at which the 225-state is favored both entropically and enthalpically, causing the Gibbs energy to disfavor the 36-state. From this, the authors concluded that HF favored the 36-phase thermodynamically, with the HF enthalpy causing its Gibbs energy to be lower without the entropy favoring either state. The authors were particularly intrigued by HF entropically favoring neither state, so they decided to further investigate why this was so.

The entropy measured in the previous data set is made up of two contributors: conformational and vibrational entropy. The conformational entropy was the same for all four hydrogen-halides as they were modeled in the same structures, and the conformational entropy for all favored the 225-structure. This could explain why the other hydrogen-halides favored the 225-state at some point, but it could not explain why HF favored neither state. Thus, the authors decided to examine the vibrational entropy to see how it impacted the system.

In the final data set chosen for the task, the authors measured the difference in vibrational entropy between the 36- and 225-state. Vibrational entropy, which is the space that atoms explore in their solid phase as they vibrate, behaved quite differently for HF compared to the other hydrogen-halides (Fultz, 2010). HF seems to exhibit significantly more vibrational entropy in the 36-structure compared to the 225, which means that it can explore more space in its 36-state compared to the 225-state. Whereas the other hydrogen-halides seemed to exhibit a preference for neither state vibrationally. From this, the authors concluded that HF's vibrational preference for the 36-state and its conformational preference for the 225-state were roughly the same magnitude and counteracted each other, which caused HF's total entropy to have a near zero value that favored neither state.

The authors argued that the lack of phase change for HF was due to enthalpy favoring the 36-state and a near-zero entropy that cannot counteract it, resulting in a negative Gibbs value favoring the 36-state. The near-zero entropy observed was caused by HF's conformational entropy and vibrational entropy canceling each other out. The authors argue that HF's vibrational entropy favors the 36-structure due to its hydrogen bonding flexing and stretching within the structure. There is a larger magnitude of hydrogen bonding present in the 36-phase for HF, which results in more flexing and stretching of the hydrogen bonds, ultimately causing the atoms to move more within their phase space.

References

Adams, A. L. (2023). *Investigating Students' Sensemaking When Engaging with Anomalous Data* [Master's Thesis, The University of Utah]. https://doi.org/10.13140/RG.2.2.22336.20486

Alicke, M. D. (2007). In Defense of Social Comparison. *Revue internationale de psychologie sociale*, *20*(1), 11–29. https://www.cairn.info/load_pdf.php?ID_ARTICLE=RIPSO_201_0011

Angra, A., & Gardner, S. M. (2017). Reflecting on graphs: Attributes of graph choice and construction practices in biology. *CBE Life Sciences Education*, *16*(3), 1–15. https://doi.org/10.1187/cbe.16-08-0245

Anker-Hansen, J., & Andrée, M. (2015). More Blessed to Give Than Receive – A Study of Peer-assessment of Experimental Design. *Procedia - Social and Behavioral Sciences*, *167*, 65–69. https://doi.org/10.1016/j.sbspro.2014.12.643

Barnard, R. A., & Shultz, G. V. (2020). "Most important is that they figure out how to solve the problem": how do advisors conceptualize and develop research autonomy in chemistry doctoral students? *Higher Education*, *79*(6), 981–999. https://doi.org/10.1007/s10734-019-00451-y

Beach, S. R. H., & Tesser, A. (2000). Self-Evaluation Maintenance and Evolution: Some Speculative Notes. In J. Suls & L. Wheeler (Eds.), *Handbook of Social Comparison*. Kluwer Academic.

Becker, N. M., Rupp, C. A., & Brandriet, A. (2017). Engaging students in analyzing and interpreting data to construct mathematical models: An analysis of students' reasoning in a method of initial rates task. *Chemistry Education Research and Practice*, *18*(4), 798–810. https://doi.org/10.1039/c6rp00205f

Berg, S. A., & Moon, A. (2022). Prompting hypothetical social comparisons to support chemistry students' data analysis and interpretations. *Chemistry Education Research and Practice*, *23*(1), 124–136. https://doi.org/10.1039/d1rp00213a

Berg, S. A., & Moon, A. (2023). A characterization of chemistry learners' engagement in data analysis and interpretation. *Chem. Educ. Res. Pract.*, *24*(1), 36–49. https://doi.org/10.1039/D2RP00154C

Bodesheim, D., Kieslich, G., Johnson, M., & Butler, K. T. (2020). Understanding the Balance of Entropy and Enthalpy in Hydrogen-Halide Noncovalent Bonding. *Journal of Physical Chemistry Letters*, *11*(9), 3495–3500. https://doi.org/10.1021/acs.jpclett.0c00817

Bolger, M. S., Osness, J. B., Gouvea, J. S., & Cooper, A. C. (2021). Supporting scientific practice through model-based inquiry: A students'-eye view of grappling with data, uncertainty, and community in a laboratory experience. *CBE Life Sciences Education*, *20*(4). https://doi.org/10.1187/cbe.21-05-0128

Brooks, J., McCluskey, S., Turley, E., & King, N. (2015). The Utility of Template Analysis in Qualitative Psychology Research. *Qualitative Research in Psychology*, *12*(2), 202–222. https://doi.org/10.1080/14780887.2014.955224

Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research*, *65*(3), 245–281. https://doi.org/10.3102/00346543065003245

Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*, *20*(2), 155–166. https://doi.org/10.1016/j.learninstruc.2009.02.014

Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment and Evaluation in Higher Education*, *43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Carmi, E., Yates, S. J., Lockley, E., & Pawluczuk, A. (2020). Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review*, *9*(2). https://doi.org/10.14763/2020.2.1481

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, *4*(2), 75–100. https://doi.org/10.1037/1076-898X.4.2.75

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

Chinn, C. A., & Brewer, W. F. (1993). The Role of Anomalous Data in Knowledge Acquisition: A Theoretical Framework and Implications for Science Instruction. *Review of Educational Research*, *63*(1), 1–49. https://doi.org/10.3102/00346543063001001

Chinn, C. A., & Brewer, W. F. (1998). An Empirical Test of a Taxonomy of Responses to Anomalous Data in Science. *Journal of Research in Science Teaching*, *35*(6), 623–654. https://doi.org/10.1002/(SICI)1098-2736(199808)35:6<623::AID-TEA3>3.0.CO;2-O

Chinn, C. A., & Brewer, W. F. (2001). Models of data: A theory of how people evaluate data. *Cognition and Instruction*, *19*(3), 323–393. https://doi.org/10.1207/S1532690XCI1903_3

Cho, K., & MacArthur, C. (2011). Learning by Reviewing. *Journal of Educational Psychology*, *103*(1), 73–84. https://doi.org/10.1037/a0021950

Connor, M. C., Glass, B. H., Finkenstaedt-Quinn, S. A., & Shultz, G. V. (2021). Developing Expertise in 1H NMR Spectral Interpretation. *Journal of Organic Chemistry*, *86*(2), 1385–1395. https://doi.org/10.1021/acs.joc.0c01398

Connor, M. C., & Shultz, G. V. (2021). Problem Solving Using NMR and IR Spectroscopy for Structural Characterization in Organic Chemistry. In G. Tsaparlis (Ed.), *Problems and Problem Solving in Chemistry Education: Analysing Data, Looking for Patterns and Making Deductions* (p. 0). The Royal Society of Chemistry. https://doi.org/10.1039/9781839163586-00166

Cooper, M. M., & Klymkowsky, M. W. (2013). Chemistry, Life, the Universe, and Everything: A New Approach to General Chemistry, and a Model for Curriculum Reform. *Journal of Chemical Education*, *90*(9), 1116–1122. https://doi.org/10.1021/ed300456y

Cooper, M. M., Caballero, M. D., Ebert-May, D., Fata-Hartley, C. L., Jardeleza, S. E., Krajcik, J. S., Laverty, J. T., Matz, R. L., Posey, L. A., & Underwood, S. M. (2015). Challenge faculty to transform STEM learning. *Science*, *350*(6258), 281 LP – 282. https://doi.org/10.1126/science.aab0933

Creswell, J. W., & Poth, C. N. (2016). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. SAGE Publications.

Crujeiras-Pérez, B., & Jiménez-Aleixandre, M. P. (2019). Students' Progression in Monitoring Anomalous Results Obtained in Inquiry-Based Laboratory Tasks. *Research in Science Education*, *49*(1), 243–264. https://doi.org/10.1007/s11165-017-9641-3

Dijkstra, P., Kuyper, H., Van Der Werf, G., Buunk, A. P., & Van Der Zee, Y. G. (2008). Social comparison in the classroom: A Review. *Review of Educational Research*, *78*(4), 828–879. https://doi.org/10.3102/0034654308321210

Doidge, E. D., Carson, I., Tasker, P. A., Ellis, R. J., Morrison, C. A., & Love, J. B. (2016). A Simple Primary Amide for the Selective Recovery of Gold from Secondary Resources. *Angewandte Chemie - International Edition*, *55*(40), 12436–12439. https://doi.org/10.1002/anie.201606113

Donkor, B., & Harshman, J. (2023). Learning Goals and Priorities Identified by an Examination of Chemistry Graduate Handbooks. *Journal of Chemical Education*, *100*(10), 3774–3783. https://doi.org/10.1021/acs.jchemed.3c00062

Duncan, R. G., Chinn, C. A., & Barzilai, S. (2018). Grasp of evidence: Problematizing and expanding the next generation science standards' conceptualization of evidence. *Journal of Research in Science Teaching*, *55*(7), 907–937. https://doi.org/10.1002/tea.21468

Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education*, *32*, 268–291. https://doi.org/10.3102/0091732X07309371

Festinger, L. (1954). A Theory of Social Comparison Processes. In *Human relations.* (Vol. 7, Issue 2, pp. 117–140). Sage Pub. https://doi.org/10.1177/001872675400700202

Finkenstaedt-Quinn, S. A., Polakowski, N., Gunderson, B., Shultz, G. V., & Gere, A. R. (2021). Utilizing Peer Review and Revision in STEM to Support the Development of Conceptual Knowledge Through Writing. *Written Communication*, 074108832110060. https://doi.org/10.1177/07410883211006038

Finkenstaedt-Quinn, S. A., Snyder-White, E. P., Connor, M. C., Gere, A. R., & Shultz, G. V. (2019). Characterizing Peer Review Comments and Revision from a Writing-to-Learn Assignment Focused on Lewis Structures. *Journal of Chemical Education*, *96*(2), 227–237. https://doi.org/10.1021/acs.jchemed.8b00711

Finkenstaedt-Quinn, S. A., Watts, F. M., & Shultz, G. V. (2024). Reading, receiving, revising: A case study on the relationship between peer review and revision in writing-to-learn. *Assessing Writing*, *59*. https://doi.org/10.1016/j.asw.2024.100808

Flower, L., Hayes, J. R., Carey, L., Schriver, K., & Stratman, J. (1986). Detection, Diagnosis, and the Strategies of Revision. *College Composition and Communication*, *37*(1), 16–55. https://doi.org/10.2307/357381

Ford, M. J. (2012). A Dialogic Account of Sense-Making in Scientific Argumentation and Reasoning. *Cognition and Instruction*, *30*(3), 207–245. https://doi.org/10.1080/07370008.2012.689383

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, *32*(2), 124–158. https://doi.org/10.2307/749671

Fultz, B. (2010). Vibrational thermodynamics of materials. *Progress in Materials Science*, *55*(4), 247–352. https://doi.org/https://doi.org/10.1016/j.pmatsci.2009.05.002

Gardner, S. K. (2007). "I heard it through the grapevine": Doctoral student socialization in chemistry and history. *Higher Education*, *54*(5), 723–740. https://doi.org/10.1007/s10734-006-9020-x

Gardner, S. K. (2008). "What's too much and what's too little?": the process of becoming an independent researcher in doctoral education. *The Journal of Higher Education*, *79*(3), 326–350.

Glazer, N. (2011). Challenges with graph interpretation: A review of the literature. *Studies in Science Education*, *47*(2), 183–210. https://doi.org/10.1080/03057267.2011.605307

González-Howard, M., & McNeill, K. L. (2020). Acting with epistemic agency: Characterizing student critique during argumentation discussions. *Science Education*, *104*(6), 953–982. https://doi.org/10.1002/sce.21592

Gouvea, J., Sawtelle, V., & Nair, A. (2019). Epistemological progress in physics and its impact on biology. *Physical Review Physics Education Research*, *15*(1), 10107. https://doi.org/10.1103/PhysRevPhysEducRes.15.010107

Greenwood, D. (2017). Social Comparison Theory. In *The International Encyclopedia of Media Effects* (pp. 1–9). https://doi.org/https://doi.org/10.1002/9781118783764.wbieme0089

Grolemund, G., & Wickham, H. (2014). A Cognitive Interpretation of Data Analysis. *International Statistical Review*, *82*(2), 184–204.

Hammer, D. (2000). Student resources for learning introductory physics. *American Journal of Physics*, *68*(S1), S52–S59. https://doi.org/10.1119/1.19520

Hammer, D., Elby, A., Scherr, R. E., Redish, E. F., Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2004). *Resources , framing , and transfer*. *Rec 0087519*, 1–26.

Heisterkamp, K., & Talanquer, V. (2015). Interpreting Data: The Hybrid Mind. *Journal of Chemical Education*, *92*(12), 1988–1995. https://doi.org/10.1021/acs.jchemed.5b00589

Hershkowitz, R., Schwarz, B. B., & Dreyfus, T. (2001). Abstraction in Context: Epistemic Actions. In *Source: Journal for Research in Mathematics Education* (Vol. 32, Issue 2).

Ion, G., Sánchez Martí, A., & Agud Morell, I. (2019). Giving or receiving feedback: which is more beneficial to students' learning? *Assessment and Evaluation in Higher Education*, *44*(1), 124–138. https://doi.org/10.1080/02602938.2018.1484881

Ivanjek, L., Susac, A., Planinic, M., Andrasevic, A., & Milin-Sipus, Z. (2016). Student reasoning about graphs in different contexts. *Physical Review Physics Education Research*, *12*(1), 1–13. https://doi.org/10.1103/PhysRevPhysEducRes.12.010106

Jeong, H., Songer, N. B., & Lee, S. Y. (2007). Evidentiary competence: Sixth graders' understanding for gathering and interpreting evidence in scientific investigations. *Research in Science Education*, *37*(1), 75–97. https://doi.org/10.1007/s11165-006-9014-9

Jiménez-Aleixandre, M. P., & Crujeiras, B. (2017). Epistemic Practices and Scientific Practices in Science Education. In K. S. Taber & B. Akpan (Eds.), *Science Education: An International Course Companion* (pp. 69–80). SensePublishers. https://doi.org/10.1007/978-94-6300-749-8_5

Kampourakis, K., & McCain, K. (2020). Uncertainty : how it makes science advance . In *Uncertainty : how it makes science advance*. Oxford University Press.

Kanari, Z., & Millar, R. (2004). Reasoning from data: How students collect and interpret data in science investigations. *Journal of Research in Science Teaching*, *41*(7), 748–769. https://doi.org/10.1002/tea.20020

Kapon, S., & diSessa, A. A. (2012). Reasoning Through Instructional Analogies. *Cognition and Instruction*, *30*(3), 261–310. https://doi.org/10.1080/07370008.2012.689385

Karch, J. M., & Sevian, H. (2022). Development of a framework to capture abstraction in physical chemistry problem solving. *Chemistry Education Research and Practice*, *23*(1), 55–77. https://doi.org/10.1039/d1rp00119a

Klein, G. a., Phillips, J. K., Rall, E. L., & Peluso, D. a. (2007). A Data-Frame Theory of Sensemaking. In *Expertise out of context* (pp. 113–155).

Klein, G., & Moon, B. (2006). Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems*, *21*(5), 88–92. https://doi.org/10.1109/MIS.2006.100

Kuhn, D., Arvidsson, T. S., Lesperance, R., & Corprew, R. (2017). Can Engaging in Science Practices Promote Deep Understanding of Them? *Science Education*, *101*(2), 232–250. https://doi.org/10.1002/sce.21263

Lai, K., Cabrera, J., Vitale, J. M., Madhok, J., Tinker, R., & Linn, M. C. (2016). Measuring Graph Comprehension, Critique, and Construction in Science. *Journal of Science Education and Technology*, *25*(4), 665–681. https://doi.org/10.1007/s10956-016-9621-9

Latour, B. (1999). *Pandora's hope: essays on the reality of science studies*. Harvard university press.

Levine, John. M. (1983). Social Comparison and Education. In M. C. Levine, J. M., Wang (Ed.), *Teacher and Student Perception: Implications for Learning* (pp. 29–55). Lawrence Earlbaum and Associates, Inc.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. SAGE Publications.

Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing*, *18*(1), 30–43. https://doi.org/10.1016/j.jslw.2008.06.002

Martin, R. (2000). “Can I Do X?”: Using the Proxy Comparison Model to Predict Performance. In J. Suls & L. Wheeler (Eds.), *Handbook of Social Comparison*. Kluwer Academic.

Masnick, A. M., & Morris, B. J. (2022). A Model of Scientific Data Reasoning. *Education Sciences*, *12*(2), 1–19. https://doi.org/10.3390/educsci12020071

Mason, M. M. (2012). Motivation, Satisfaction, and Innate Psychological Needs. *International Journal of Doctoral Studies*, *7*, 259–273.

May, J. M., Barth-Cohen, L. A., Gerton, J. M., De Grandi, C., & Adams, A. L. (2022). Student sensemaking about inconsistencies in a reform-based introductory physics lab. *Physical Review Physics Education Research*, *18*(2). https://doi.org/10.1103/PhysRevPhysEducRes.18.020134

May, J. M., Barth-Cohen, L. A., Gerton, J. M., & Grandi, C. De. (2020). Students’ dynamic engagement with experimental data in a physics laboratory setting. *Physics Education Research Conference Proceedings*, 315–320. https://doi.org/10.1119/perc.2020.pr.May

McConlogue, T. (2015). Making judgements: investigating the process of composing and receiving peer feedback. *Studies in Higher Education*, *40*(9), 1495–1506. https://doi.org/10.1080/03075079.2013.868878

Meister, S., Krell, M., Göhner, M., & Upmeier zu Belzen, A. (2021). Pre-service Biology Teachers’ Responses to First-Hand Anomalous Data During Modelling Processes. *Research in Science Education*, *51*(6), 1459–1479. https://doi.org/10.1007/s11165-020-09929-7

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative Research : a Guide to Design and Implementation* (Fourth edition). Jossey-Bass, a Wiley brand.

Miles, M. B., Michael Huberman, A., & Saldaña, J. (2014). *Qualitative Data Analysis: A Methods Sourcebook* (Third edit). SAGE.

Miller, M. K., Reichert, J., & Flores, D. (2015). Social Comparison. In *The Blackwell Encyclopedia of Sociology*. https://doi.org/https://doi.org/10.1002/9781405165518.wbeoss140.pub2

Moon, A., Stanford, C., Cole, R., & Towns, M. (2017). Decentering: A Characteristic of Effective Student–Student Discourse in Inquiry-Oriented Physical Chemistry Classrooms. *Journal of Chemical Education*, *94*(7), 829–836. https://doi.org/10.1021/acs.jchemed.6b00856

National Academies of Sciences, E. and M. (2018). *Graduate STEM Education for the 21st Century* (A. Leshner & L. Scherer, Eds.). The National Academies Press. https://doi.org/10.17226/25038

National Research Council. (2012). A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. In *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. The National Academies Press. https://doi.org/10.17226/13165

Nelson, J. (2017). Using conceptual depth criteria: addressing the challenge of reaching saturation in qualitative research. *Qualitative Research*, *17*(5), 554–570. https://doi.org/10.1177/1468794116679873

Nicol, D. (2020). The power of internal feedback: exploiting natural comparison processes. *Assessment and Evaluation in Higher Education*, *0*(0), 1–23. https://doi.org/10.1080/02602938.2020.1823314

Nicol, D., & Kushwah, L. (2023). Shifting feedback agency to students by having them write their own feedback comments. *Assessment and Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2023.2265080

Nicol, D., & McCallum, S. (2021). Making internal feedback explicit: exploiting the multiple comparisons that occur during peer review. *Assessment and Evaluation in Higher Education*, *0*(0), 1–19. https://doi.org/10.1080/02602938.2021.1924620

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assessment and Evaluation in Higher Education*, *39*(1), 102–122. https://doi.org/10.1080/02602938.2013.795518

Osborne, J. F., Henderson, J. B., MacPherson, A., Szu, E., Wild, A., & Yao, S. Y. (2016). The development and validation of a learning progression for argumentation in science. *Journal of Research in Science Teaching*, *53*(6), 821–846. https://doi.org/10.1002/tea.21316

Patchan, M. M., & Schunn, C. D. (2015). Understanding the benefits of providing peer feedback: how students respond to peers' texts of varying quality. *Instructional Science*, *43*(5), 591–614. https://doi.org/10.1007/s11251-015-9353-x

Pepitone, E. A. (1972). Comparison Behavior in Elementary School Children. *American Educational Research Journal*, *9*(1), 45–63. https://doi.org/10.3102/00028312009001045

Phillips, A. M. L., Sundstrom, M., Wu, D. G., & Holmes, N. G. (2021). Not engaging with problems in the lab: Students' navigation of conflicting data and models. *Physical Review Physics Education Research*, *17*(2). https://doi.org/10.1103/PhysRevPhysEducRes.17.020112

Phillips, A. M. L., Watkins, J., & Hammer, D. (2018). Beyond "asking questions": Problematizing as a disciplinary activity. *Journal of Research in Science Teaching*, *55*(7), 982–998. https://doi.org/10.1002/tea.21477

Piaget, J. (1955). *The Language and Thought of the Child*. Meridian Books.

Pinker, S., & Feedle, R. (1990). A theory of graph comprehension. In *Artificial Intelligence and the Future of Testing* (pp. 73–126).

Pomery, E. A., Gibbons, F. X., & Stock, M. L. (2012). Social Comparison. *Encyclopedia of Human Behavior: Second Edition*, 463–469. https://doi.org/10.1016/B978-0-12-375000-6.00332-3

Potgieter, M., Harding, A., & Engelbrecht, J. (2008). Transfer of Algebraic and Graphical Thinking between Mathematics and Chemistry. *Journal of Research in Science Teaching*, *45*(2), 196–218. https://doi.org/10.1002/tea

Ratwani, R. M., Trafton, J. G., & Boehm-Davis, D. A. (2008). Thinking graphically: Connecting vision and cognition during graph comprehension. *Journal of Experimental Psychology: Applied*, *14*(1), 36–49. https://doi.org/10.1037/1076-898X.14.1.36

Roth, W., & Bowen, G. M. (2000). Learning Difficulties Related to Graphing : A Hermeneutic. *Research in Science Education*, *30*(1), 123–139.

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment and Evaluation in Higher Education*, *35*(5), 535–550. https://doi.org/10.1080/02602930903541015

Shah, P., & Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. In *Journal of Experimental Psychology: General* (Vol. 124, Issue 1, pp. 43–61). American Psychological Association. https://doi.org/10.1037/0096-3445.124.1.43

Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, *14*(1), 47–69. https://doi.org/10.1023/A:1013180410169

Sharon, A. J., & Baram-Tsabari, A. (2020). Can science literacy help individuals identify misinformation in everyday life? *Science Education*, *104*(5), 873–894. https://doi.org/10.1002/sce.21581

Singer, S. R., Nielsen, N. R., & Schweingruber, H. A. (2012). *Discipline-based education research*. National Academies Press.

Slominski, T., Fugleberg, A., Christensen, W. M., Buncher, J. B., & Momsen, J. L. (2020). Using framing as a lens to understand context effects on expert reasoning. *CBE Life Sciences Education*, *19*(3), 1–15. https://doi.org/10.1187/cbe.19-11-0230

Smith, W. P., & Arnkelsson, G. B. (2000). Stability of Related Attributes and the Inference of Ability through Social Comparison. In J. Suls & L. Wheeler (Eds.), *Handbook of Social Comparison*. Kluwer Academic.

Taber, K. S. (2017). Reflecting the Nature of Science in Science Education. In K. S. Taber & B. Akpan (Eds.), *Science Education: An International Course Companion* (pp. 23–37). SensePublishers. https://doi.org/10.1007/978-94-6300-749-8_2

Talanquer, V., & Pollard, J. (2010). Let's teach how we think instead of what we know. *Chemistry Education Research and Practice*, *11*(2), 74–83. https://doi.org/10.1039/c005349j

Teuscher, D., Moore, K. C., & Carlson, M. P. (2016). Decentering: A construct to analyze and explain teacher actions as they relate to student thinking. *Journal of Mathematics Teacher Education*, *19*(5), 433–456. https://doi.org/10.1007/s10857-015-9304-0

The American Chemical Society. (2012). *Advancing graduate education in the chemical sciences*.

Urbanek, M. T., Moritz, B., & Moon, A. (2023). Exploring students' dominant approaches to handling epistemic uncertainty when engaging in argument from evidence. *Chemistry Education Research and Practice*. https://doi.org/10.1039/D3RP00035D

van Popta, E., Kral, M., Camp, G., Martens, R. L., & Simons, P. R. J. (2017). Exploring the value of peer feedback in online learning for the provider. *Educational Research Review*, *20*, 24–34. https://doi.org/10.1016/j.edurev.2016.10.003

Watts, F. M., Finkenstaedt-Quinn, S. A., & Shultz, G. V. (2024). Examining the role of assignment design and peer review on student responses and revisions to an organic chemistry writing-to-learn assignment. *Chemistry Education Research and Practice*. https://doi.org/10.1039/D4RP00024B

Yan, Z., & Brown, G. T. L. (2017). A cyclical self-assessment process: towards a model of how students engage in self-assessment. *Assessment and Evaluation in Higher Education*, *42*(8), 1247–1262. https://doi.org/10.1080/02602938.2016.1260091

Zagallo, P., Meddleton, S., & Bolger, M. S. (2016). Teaching real data interpretation with models (TRIM): Analysis of student dialogue in a large-enrollment cell and developmental biology course. *CBE Life Sciences Education*, *15*(2), 1–18. https://doi.org/10.1187/cbe.15-11-0239

Zhou, J., & Moon, A. (2023). "To Be Honest, I Didn't Even Use the Data": Organic Chemistry Students' Engagement in Data Analysis and Interpretation. *Journal of Chemical Education*, *100*, 80–90. https://doi.org/10.1021/acs.jchemed.2c00840