

DISSERTATION

I. COGNITIVE AND INSTRUCTIONAL FACTORS RELATING TO
STUDENTS' DEVELOPMENT OF PERSONAL MODELS OF CHEMICAL SYSTEMS
IN THE GENERAL CHEMISTRY LABORATORY
II. SOLVATION IN SUPERCRITICAL CARBON DIOXIDE/ETHANOL MIXTURES
STUDIED BY MOLECULAR DYNAMICS SIMULATION

Submitted by

Seth Anthony

Department of Chemistry

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2014

Doctoral Committee:

Advisor: Dawn Rickey

Co-Advisor: Branka M. Ladanyi

Grzegorz Szamel

Amy L. Prieto

Edward L. DeLosh

Copyright by Seth Anthony 2014

All Rights Reserved

ABSTRACT

I. COGNITIVE AND INSTRUCTIONAL FACTORS RELATING TO STUDENTS' DEVELOPMENT OF PERSONAL MODELS OF CHEMICAL SYSTEMS IN THE GENERAL CHEMISTRY LABORATORY

II. SOLVATION IN SUPERCRITICAL CARBON DIOXIDE/ETHANOL MIXTURES STUDIED BY MOLECULAR DYNAMICS SIMULATION

Part I. Students' participation in inquiry-based chemistry laboratory curricula, and, in particular, engagement with key thinking processes in conjunction with these experiences, is linked with success at the difficult task of "transfer" – applying their knowledge in new contexts to solve unfamiliar types of problems. We investigate factors related to classroom experiences, student metacognition, and instructor feedback that may affect students' engagement in key aspects of the Model-Observe-Reflect-Explain (MORE) laboratory curriculum – production of written molecular-level models of chemical systems, describing changes to those models, and supporting those changes with reference to experimental evidence – and related behaviors. Participation in introductory activities that emphasize reviewing and critiquing of sample models and peers' models are associated with improvement in several of these key aspects. When students' self-assessments of the quality of aspects of their models are solicited, students are generally overconfident in the quality of their models, but these self-ratings are also sensitive to the strictness of grades assigned by their instructor. Furthermore, students who produce higher-quality models are also more accurate in their self-assessments, suggesting the importance of self-evaluation as part of the model-writing process. While

the written feedback delivered by instructors did not have significant impacts on student model quality or self-assessments, students' resubmissions of models were significantly improved when students received "reflective" feedback prompting them to self-evaluate the quality of their models. Analysis of several case studies indicates that the content and extent of molecular-level ideas expressed in students' models are linked with the depth of discussion and content of discussion that occurred during the laboratory period, with ideas developed or personally committed to by students during the laboratory period being likely to appear in students' post-laboratory refined models. These discussions during the laboratory period are primarily prompted by factors external to the students or their laboratory groups such as questions posed by the instructor or laboratory materials.

Part II. Solvation of polar molecules within non-polar supercritical carbon dioxide is often facilitated by the introduction of polar cosolvents as entrainers, which are believed to preferentially surround solute molecules. Molecular dynamics simulations of supercritical carbon dioxide/ethanol mixtures reveal that ethanol molecules form hydrogen-bonded aggregates of varying sizes and structures, with cyclic tetramers and pentamers being unusually prevalent. The dynamics of ethanol molecules within these mixtures at a range of thermodynamic conditions can largely be explained by differences in size and structure in these aggregates. Simulations that include solute molecules reveal enhancement of the polar cosolvent around hydrogen-bonding sites on the solute molecules, corroborating and helping to explain previously reported experimental trends in solute mobility.

ACKNOWLEDGEMENTS

I began my journey towards a Ph.D. emphasizing chemistry education in order to learn what made science instruction truly effective, to contribute in a small way to that body of knowledge, and, most importantly, to learn how to become the most effective chemistry teacher that I could be. When I look back on the journey that has led me here, I'm reminded of the many individuals who have encouraged, supported, challenged, and guided me down this path, and I wish to acknowledge just a few of the most important here.

My journey towards being a scientist and a teacher was sparked in so many ways by my grandmother, Betty Dilday, who bought me my very first chemistry set when I was in the second grade and who listened – patiently and with genuine interest – to many of my early attempts at teaching – explaining the internet, sailboats, or black holes. Her encouragement of my curiosity and natural interest in explaining the world – to myself or to others – has truly set me off on my path in life.

I would not be where I am today without having been pushed by amazing teachers during my entire academic career. When I think about my love of science and mathematics, I cannot help but think back all the way to elementary school and the classrooms of Ms. Carol Nebrat and Ms. Phyllis Beeson, who not only encouraged my enthusiasm, but encouraged me to share my discoveries and joy in science and math with others. In college at North Carolina State University, I was mentored and encouraged by a number of professors who pushed me in the direction that I finally took. This included Dr. Kay Sandberg, who first got me involved in undergraduate research; Dr. Brent Gunnoe, in whose lab I learned that I am in no sense a synthetic

chemist; Dr. Jonathan Kramer, who encouraged me to think outside the box in pursuing my academic interests; and Dr. Maria Oliver-Hoyo, without whom I wouldn't have even known chemistry education existed as a field of study. I would be remiss if I neglected to mention the late Dr. J.R. Boyd, who taught at the Governors' School of North Carolina, and who demonstrated by example how non-traditional teaching methods could engage talented students. Foremost among the teachers who have inspired me, though, is Dr. Forrest Hentz, in whose general chemistry class I saw firsthand the joys of learning and teaching chemistry, and who remains my role model for inspiring and engaging students whenever I step into a classroom.

Once I reached Colorado State University for graduate work, my chemistry education research wouldn't have been possible without the support of Colorado State University's chemistry department for innovations within the curriculum – for this, I want to thank the department chairs during my time here, Dr. Tony Rappe and Dr. Ellen Fisher. I also wish to thank Dr. Nancy Boldt, the general chemistry laboratory coordinator, and all of the staff on the 4th floor of Yates Hall, for understanding as I barged into and out of dozens of chemistry lab classrooms with video cameras and stacks of lab reports to scan.

My graduate advisors– Dr. Dawn Rickey and Dr. Branka Ladanyi – have been amazingly supportive, demanding high-quality work from me while expressing confidence which I did not always have in my ability to complete it. I especially want to thank them for their patience and understanding as I sought to juggle two essentially unrelated projects – one in chemistry education, one in physical chemistry – as part of my graduate program, and as completing my graduate program took longer than I anticipated when started a faculty job while not yet done writing this dissertation. I

must also acknowledge Dr. Ken Usher, my department chair in my new job at Oregon Institute of Technology, for his understanding and gentle encouragement towards finishing this dissertation.

I am also grateful to faculty both within Colorado State University's chemistry department and outside it who have encouraged and supported me in my graduate career and offered helpful advice, particularly Dr. Amy Prieto, Dr. Nancy Levinger, Dr. Matt Rhodes, Dr. Lisa Dysleski, Dr. Melonie Teichert, and Dr. Lydia Tien. Without their guidance, mentorship, and encouraging conversations, my graduate program would have looked much different. To my colleagues and office-mates in the chemistry department – particularly Laura Wally, Colin Blair, and Ryan Trott in the chemistry education group, and Dr. Janamajeya Chowdhary and Dr. Anatoli Milischuk in the physical chemistry group, I offer sincere thanks for sharing laughter, frustration, and meaningful conversations about research and the process of research.

My family has always shown tremendous confidence in me, and I can only hope that I can continue to live up to their expectations. The lessons I've learned from them are too numerous to list, but I must acknowledge my father, who taught me to always check my assumptions, and my mother, who taught me to always persist through difficult situations.

Most importantly, I am unendingly grateful to my wife, Eve Klopff, who has shared this journey towards a Ph.D. alongside me. Life is so much richer when you have someone to share both joys and frustrations with, and I do not believe I would have been able to reach this finish line without her support, encouragement, and unfailing confidence

Thank you, all, for your support on this journey.

DEDICATION

*This dissertation is dedicated
to my grandmother, Betty Dilday,
who began me on my journey
to learn, to teach, and to try
to make the world a better place for others;
and to my son, Ian Matthias Klopff,
who will continue that journey after me.*

TABLE OF CONTENTS

Abstract	ii
Acknowledgments	iv
Dedication	vii
PART I: COGNITIVE AND INSTRUCTIONAL FACTORS RELATING TO STUDENTS' DEVELOPMENT OF PERSONAL MODELS OF CHEMICAL SYSTEMS IN THE GENERAL CHEMISTRY LABORATORY	1
Chapter 1: Introduction: Motivating Questions and the Model-Observe-Reflect-Explain Thinking Frame	2
1.1 – The MORE Thinking Frame	5
1.2 – Dissertation Structure	7
Chapter 2: Development of a Model-Observe-Reflect-Explain Laboratory Activity and the Effects of Context	10
2.1 – Precipitation Reactions and Conductivity	11
2.2 – Experimental Structure	13
2.2.1 – Activities Preceding Precipitation Experiments	13
2.2.2 – Initial Model Assignment	15
2.2.3 – Part I: Lead Iodide	16
2.2.4 – Part II: Sulfates and Hydroxides	19
2.2.5 – Part III: Student-Designed Experiments	21
2.3 – Outcomes: Refined Models	22

2.4 – Effects of Context	24
2.4.1 – Conductivity Context	24
2.4.2 – Precipitation Reaction Context	29
2.5 – Discussion	32

Chapter 3: Investigating the Use of Modeling and Critiquing Activities to

Enhance Student Engagement in Scientific Thinking Processes	35
3.1 – <i>Learning to Construct and Refine Models</i> Activities	36
3.2 – Participants and Methods.....	41
3.3 – Molecular-Level Language	43
3.4 – Reflection on Model Revisions and Use of Evidence	45
3.5 – Outcomes: Model Quality for 2006 and 2008 Cohorts	49
3.6 – Outcomes: Model Quality for 2009 Cohort	49
3.7 – Discussion	50

Chapter 4: Student Self-Assessments and Relationships with

Model Quality and Instructional Conditions	53
4.1 – Background	54
4.2 – Study Population and Context.....	57
4.3 – “Expert” Model Quality Ratings	63
4.4 – The InSPRE Activity	74
4.5 – Analysis of Student Self-Assessments	77
4.6 – Calibration of Student Self-Assessments	89
4.7 – Discussion	111

Chapter 5: Relationships between Written Instructor Feedback, Model Quality and Self-assessment, and Model Resubmissions	113
5.1 – Study Design.....	116
5.2 – Analysis of Written Feedback Delivered	119
5.3 – Comparisons Between Feedback Groups	125
5.4 – Analysis by Feedback Received	132
5.5 – Potential Correlates with End-of-Semester Measures	139
5.6 – Analysis of Resubmissions.....	145
5.7 – Discussion	157
 Chapter 6: Analysis of Student Reflection From Video.....	159
6.1 – Selection of Video Groups and Data Collection.....	160
6.2 – Factors Instigating Molecular-Level Reflections	163
6.3 – Case Studies of Student Discussions	167
6.3.1 – Case Study 1: “The High Lattice Theory”	168
6.3.2 – Case Study 2: “Anyone... Remember?”	180
 Chapter 7: Further Research Questions	188
 References: Part I.....	194

PART II: SOLVATION IN SUPERCRITICAL CARBON DIOXIDE/ETHANOL

MIXTURES STUDIED BY MOLECULAR DYNAMICS SIMULATION 211

Chapter 8: Introduction: Solvation in Supercritical

Carbon Dioxide/Cosolvent Systems..... 212

8.1 – Previous Studies of sc-CO₂/Alcohol Systems215

8.2 – Central Questions and Dissertation Structure..... 222

Chapter 9: Structure and Dynamics of Ethanol Aggregates in

Supercritical Carbon Dioxide/Ethanol Mixtures 225

9.1 – Simulation Parameters and Conditions..... 228

9.2 – Structure of CO₂/Ethanol Mixtures..... 231

9.2.1 – Distribution Functions..... 231

9.2.2 – Hydrogen Bonding Patterns..... 236

9.3 – Dynamical Properties 241

9.3.1 – Diffusion..... 241

9.3.2 – Orientational Time Corrections 247

9.3.3 – Hydrogen Bond Dynamics 254

9.4 – Discussion 259

Chapter 10: Intermolecular Interactions Around Solutes in Supercritical Carbon Dioxide/Ethanol Mixtures	260
10.1 – Simulation Parameters and Conditions	261
10.2 – Pair Correlation Functions	266
10.3 – First Solvation Shell	271
10.4 – Solvation Shell Residence Times	275
10.5 – Ethanol/Solute Hydrogen Bonding	277
10.6 – Diffusion	279
10.7 – Discussion	282
 Chapter 11: Further Research Questions	 284
 References: Part II	 287
 APPENDIX A: FULL STUDENT TEXT OF “WHAT HAPPENS WHEN CHEMICAL COMPOUNDS ARE ADDED TO WATER” (WEEK III) LABORATORY ACTIVITY	 296
 APPENDIX B: INSTRUCTOR GUIDE FOR “WHAT HAPPENS WHEN CHEMICAL COMPOUNDS ARE ADDED TO WATER” (WEEK III) LABORATORY ACTIVITY	 305

APPENDIX C: FULL STUDENT TEXT OF “LEARNING TO CONSTRUCT AND REFINE MODELS” INTRODUCTORY ACTIVITY	313
APPENDIX D: INSTRUCTOR GUIDE FOR “LEARNING TO CONSTRUCT AND REFINE MODELS” INTRODUCTORY ACTIVITY	341
APPENDIX E: STUDENT TEXT FOR SECOND WEEK OF “LEARNING TO CONSTRUCT AND REFINE MODELS” INTRODUCTORY ACTIVITY WITH INSTRUCTOR GUIDE TEXT	361
APPENDIX F: TEXT OF ALL INITIAL AND REFINED MODEL ASSIGNMENTS FROM FALL 2008 IMPLEMENTATION	373
APPENDIX G: SAMPLE REVIEWING PACKET FROM InSPRE (IN-CLASS SELF- AND PEER-REVIEWING EXERCISE) IN-CLASS ACTIVITY	386

PART I

COGNITIVE AND INSTRUCTIONAL FACTORS RELATING TO
STUDENTS' DEVELOPMENT OF PERSONAL MODELS OF CHEMICAL SYSTEMS
IN THE GENERAL CHEMISTRY LABORATORY

Chapter 1: Introduction: Motivating Questions and the Model-Observe-Reflect-Explain (MORE) Thinking Frame

A myriad of chemistry instructors have lamented that many students still seem to possess a shallow understanding of chemistry concepts, even after determined attempts to impart knowledge in the classroom or teaching laboratory. Research literature in the field of chemistry education validates these anecdotal impressions. Students may be adept, for instance at applying formulas to solve problems and obtain correct quantitative or numerical answers while demonstrating poor conceptual knowledge of the same topic¹ – failing to correctly reason about the behavior of the molecules, atoms, ions, and other species that make up the particulate world of chemistry. Numerous studies have documented incorrect ideas or misconceptions that students hold about chemical concepts, even after instruction.²

Novel instructional approaches have been proposed to ameliorate these difficulties, and many have been tested with varying degrees of success. Efforts in laboratory curriculum development in recent decades have emphasized a shift away from the “traditional” approach, sometimes pejoratively referred to as a “verification” or “cookbook” approach, in which students follow set directions to generate data that is anticipated.³ In traditional laboratory experiments, students’ objectives are typically to verify previously known key relationships, principles, or chemical behaviors. By contrast, within inquiry-based approaches, as described by Domin (2009), students draw upon experimental evidence to inductively develop concepts and ideas that are new to them, “leading to the construction of new understanding that can be

generalized.”⁴. These approaches typically have students investigate systems or take measurements for which they do not know the outcome in advance. They may provide students with procedures and ask students to decide how to interpret the experimental evidence they encounter (“guided inquiry”), rather than having interpretation methodologies provided for them, or have students pose their own questions and design their own experiments rather than having all of the procedures and questions laid out for them in advance (“open inquiry”).⁵

Such approaches tend to make *constructivist* assumptions about the nature of learning – that knowledge arises in the mind of learners not merely due to direct transmission from an instructor, but because they explicitly create knowledge for themselves, in part by developing linkages with existing knowledge. As described by Piaget, a learner exists “in a universe that he has gradually constructed for himself;”⁶ Bodner summarizes the model with the statement: “Knowledge is constructed in the mind of the learner.”⁷

Although constructivist assumptions have informed a great deal of current science pedagogy,⁸ including inquiry-based laboratory instruction in chemistry^{5a, 5d} the specific design of aspects of these curricula or their implementation is often based on personal experience. Instructors sometimes make important decisions that affect student learning – how to introduce inquiry, how to grade assignments, what types of feedback to give, what sorts of conversations to have with students – based less on data-driven analysis and more on anecdotal personal experience.

However, even “smaller” decisions about the design and implementation of these inquiry-based curricula can be informed by research data. Rickey, Teichert, and Tien⁹ offer an example of this by contrasting students’ success at developing personal models

of chemical system after a procedurally complex “Synthesis of an Alum” laboratory experiment as compared with a simpler “Blue Dye” experiment. They describe how fewer students generated high-quality molecular-level models when they were cognitively taxed by the procedures in more complex experiment, and corroborated this conclusion with observations of interactions between student and instructors during the laboratory period. Investigation of student performance during instruction can provide general principles and insights into their implementation and design..

Studies of the potential benefits of inquiry-based instruction have also often placed inquiry-based curricula “head-to-head” with more traditional approaches. While this research is important, and helps demonstrate the effectiveness of such approaches in a quasi-controlled experimental setting, it offers limited insight into *how* to implement or design such instruction, or of why, cognitively, such approaches offer benefits for students. To deepen our understanding of why inquiry-based approaches to laboratory instruction are effective and how to implement them most effectively in real-world settings, it is necessary to examine the relationships between student and instructor actions and cognition throughout these forms of instruction and student outcomes. At its core, this is the fundamental question that guides the research questions posed in chapters 3-6 of this dissertation:

What factors – both in terms of instructional conditions and student cognition – are associated with the quality of key aspects students’ written models within a inquiry-based general chemistry laboratory course employing the Model-Observe-Reflect-Explain Thinking Frame?

In particular, this work investigates a number of factors that may influence students' performance at the complex scientific reasoning tasks demanded. Among the questions examined are:

- What are the impacts of different approaches of introducing students to inquiry-based instruction? How can introductory activities that involve critiquing and evaluation of sample and peer work facilitate students' own construction of personal models?
- What are the relationships between students' performance within the written work involved in inquiry-based laboratories and their ability to assess the quality of their own attempts at this work? How might both students' performance and self-assessment within these laboratories be influenced by the grade and written feedback provided by their instructors?
- How do the in-person interactions that take place during the laboratory period shape the development of students' personal models of chemical systems? What sorts of interactions are the most effective at promoting students' development of and reflection on molecular-level models grounded in experimental evidence?

1.1 The MORE Thinking Frame

The particular inquiry-based approach that forms the context for this work is the Model-Observe-Reflect-Explain (MORE) Thinking Frame, which was originally designed by a team at the University of California-Berkeley and has been shown to bring about many of the benefits often sought from inquiry-based instructional approaches.

Students who participate in laboratory curricula employing the MORE Thinking Frame exhibit: ⁹⁻¹⁰

- higher performance on solving examination questions, particularly those involving transfer of one knowledge out of familiar contexts to solve new types of problems not previously encountered;
- greater engagement in metacognition, or thinking about one's own thinking, particularly concerning students' personal models or understandings of chemical systems;
- more sophisticated beliefs about the nature of scientific inquiry; and,
- enhanced ability to evaluate evidence and design experiments.

The first item on this list – facilitating the ability of students to apply their ideas, personal models, or understandings in new contexts, or to solve types of problems that they have not previously encountered, is a capacity often referred to as “transfer,” and is considered a hallmark of “deep” understanding. ¹¹ It has been found that students who engage in several key thinking processes within their written work in Model-Observe-Reflect-Explain laboratory experiments show greater success at applying those models to correctly reason about new situations.

The approach helps students emulate the thinking processes of scientists by scaffolding their *thinking* about experimental evidence and its relationship with their own personal understanding of a system. This contrasts with a more traditional approach to laboratory curriculum design, which tends to focus on procedural instruction. As summarized by Rickey et al.: “Students are typically given experimental protocols telling them *what to do* in the lab. It is far less common for students to be scaffolded in *how to think* during the inquiry process” [emphasis in original].⁹

When participating in MORE laboratory experiments, students begin by describing, using words and/or pictures, their *initial models*– their understanding of a chemical system on both the macroscopic and molecular-levels (model). They then conduct experiments and record findings and measurements regarding that system and related systems (observe) while considering how these observations inform their initial models (reflect). Finally, students revisit and revise their initial models, changing them to be consistent with experimental evidence and describing – by referring to experimental evidence – how and why various aspects of their models have either changed or remained the same, in a *refined model* (explain).

Because the pre-laboratory *initial models* and post-laboratory *refined models* both include students' descriptions of their own understandings, these can be examined. Students who participate in MORE laboratory modules exhibit more scientifically correct understandings in their refined models than in their initial models.¹² Furthermore, students who produce refined models that are consistent with experimental evidence, explain correctly how their models have changed in response to that evidence, and cite specific experimental observations to justify those model refinements are seen to successfully apply those models to answer questions in new contexts – to exhibit *transfer* of their understanding.¹³ The importance of these thinking processes informs many of the examinations in this dissertation.

1.2 Dissertation Structure

In Chapter 2, I begin by introducing the MORE Thinking Frame and a key issue associated with students' understanding of chemistry – the context-dependence of students' molecular level ideas -- by illustrating the development and implementation of

a MORE laboratory experiment on the topic of precipitation reactions. This experiment employs evidence including precipitate quantity and appearance as well as conductivity measurements of the supernatant liquid to address students' molecular-level models of reactions. Analysis of students' written models from this experiment shows that, in their refined models written after the experiment, students produce more scientifically correct explanations.

This laboratory experiment builds upon a previous laboratory module in which students use conductivity as a tool to explore properties of aqueous solutions and mixtures. Because the final model from this previous module is conjoined with the initial model, the ideas expressed in the two portions of the assignment can be compared. Interestingly, students are found to exhibit contradictory views within the same assignment – a phenomenon that was also seen in prior work. An additional experiment described in this chapter probed the effect of the particular context – in this case, conductivity – on students' ideas.

Findings that students' ideas are not always activated consistently in different contexts – in other words, that students do not consistently engage in transfer– motivates work to improve students' performance on aspects of models known to be associated with transfer success. Chapters 3 through 6 focus on efforts to address the central questions of this dissertation – to understand how cognitive factors and instructional conditions relate to students' success at engaging in these key tasks within their models.

Chapter 3 describes and evaluates an introductory activity that has been implemented to help introduce students to the expectations and requirements of model-writing within MORE Thinking Frame instruction. These activities, implemented at the

start of the academic semester, have students evaluate and critique sample models provided by their instructor, models written by peers, and their own models, while leading students through a cycle of the MORE Thinking Frame. As the modeling and critiquing components of these introductory activities were expanded over successive years, students demonstrated greater success at constructing models that contained truly molecular-level explanations and higher-quality models.

Participation in model-reviewing activities may foster students' ability to self-assess their own work, which is the focus of Chapter 4. Drawing upon work in cognitive psychology showing that students are often overconfident in their own performance, I collected students' quantitative self-assessments of the quality of their own models throughout an academic semester. By comparing students' self-assessments with more "expert" ratings of their models – either by their instructor or by a researcher – we could examine their accuracy. We found that students' self-assessments or their accuracy differed between different model aspects, between students of different instructors, and between laboratory modules across the semester. Most importantly, however, it was found that the students who evaluated themselves most accurately also produced the highest quality models.

Having identified that student self-assessment accuracy is linked to model quality, Chapter 5 describes an attempt to influence the accuracy of students' self-assessments by having instructors provide written feedback specifically intended to target students' metacognitive monitoring of their own performance at model-writing. While one population of students received primarily feedback describing deficiencies in their models ("direct" feedback), another population of students received more feedback prompting them to search their own models for deficiencies ("reflective" feedback).

This experiment was conducted with a semi-controlled design, however, even when factoring in analysis of the actual feedback delivered by instructors, no relationships were found between the types of written feedback received and either the quality of models produced or the accuracy of students' self-assessments.

However, opportunities that were offered during the study for students to resubmit graded models and improve those models based on feedback did uncover differences – students who received reflective feedback on particular model aspects were more likely to make higher-quality revisions to their models than those who received feedback directly identifying their errors. Although this was not part of the original intent of the study, it does suggest some positive impacts of feedback targeted at self-monitoring.

Finally, in order to investigate relationships between in-class student discussions and student model-writing, Chapter 6 examines several case studies of groups of students conducting laboratory experiments. In particular, students appear to draw heavily on their in-class discussions to develop their models; students who originate or develop key ideas within group discussion are seen to incorporate those into their models. Moreover, discussions about the molecular-level behavior of systems or about whether students' personal models or expectations are confirmed or refuted (related to the metacognition targeted in written models) are relatively rare, even among the groups that engage in this the most, and are typically prompted by factors such as the instructor, a laboratory manual question, or an unexpected experimental observation.

Chapter 7 offers discussion of the relevance of the findings with respect to the broader literature, discusses implications and suggestions for instructors and curriculum developers, and poses additional questions for future study.

Chapter 2: Development of a Model-Observe-Reflect-Explain Laboratory Activity and the Effects of Context

In this chapter, I first describe the development of a laboratory experiment using the Model-Observe-Reflect-Explain Thinking Frame. This experiment, designed to address students' models of precipitation reactions, builds upon a previously developed MORE laboratory module, "What Happens When Chemical Compounds Are Added to Water?" Analysis of the ideas students express in their initial and refined models, and in interviews on related systems, reveal relationships between the ideas that students express and the contexts in which they express them. Students' tendencies to express different and often contradictory molecular-level ideas about different molecular-level systems offer insight into the structure of knowledge in the mind and illustrate the need for deeper understanding of how to best implement inquiry-based laboratory curricula.

2.1 Precipitation Reactions and Conductivity

Precipitation reactions – those in which the mixture of two solutions of soluble compounds results in the formation of a solid – are one of the major classes of reactions addressed in the general chemistry curriculum¹⁴, and the chemistry education literature has documented a number of scientifically inaccurate beliefs that students hold about this class of reactions.¹⁵

A detailed study by Kelly, Barrera, and Mohamed^{15b} reported that, after classroom instruction on aqueous solutions and precipitation reactions in a college chemistry course, most students interviewed still showed substantial misconceptions. When asked to produce molecular-level representations of a precipitation reaction, half

or more of students represented aqueous reactants or products as molecular species or as covalently bonded ions. In many cases, students described ionic reactants as only “breaking apart” into separate ions when solutions are mixed, then form new molecular species in the reaction mixture, either in solution or as a precipitate.

A number of laboratory experiments for the general chemistry laboratory involving precipitation reactions have been reported, using precipitation reactions to investigate reaction stoichiometry,¹⁶ precipitate identity,¹⁷ solubility product constants,¹⁸ or reaction equilibrium¹⁹. These experiments, though, tend to focus on quantification or analysis of the solid precipitate that is formed, and tend not to examine the remaining solution. At least one experiment has been reported that uses conductivity as a tool to examine the aqueous phase substances surrounding a precipitation reaction²⁰, but it was not designed to direct students to reflect on the molecular-level situation which results in their observations.

Conductivity measurements are easily performed in the general chemistry lab using commercially available equipment including computerized “probeware” systems,²¹ or using a “homemade” conductivity apparatus.²² The laboratory experiment described in this chapter has been conducted using both types of conductivity testers to examine the concentration of dissolved ions. Since the conductivity of a dilute aqueous solution increases with increasing concentration of dissolved ions, conductivity measurements can be used as a window into the relative concentrations of dissolved ions in both reactant and product mixtures.

Electrical conductivity of solutions has also been similarly employed as a tool to explore the behavior of a number of different systems, directly, and indirectly, through its use as an indicator of ion concentration. In this way, various laboratory experiments

have used conductivity as a tool to explore acid-base titrations,²³ base strength,²⁴ the critical micelle concentration of ionic surfactants,²⁵ and reaction kinetics, particularly as a means towards describing reaction mechanisms.^{22a, 26} Conductimetric titrations have also been employed in experiments for quantitative analysis,²⁷ and conductivity can even be used as a probe of ion concentration in non-aqueous systems.^{22a, 26c, 26d, 28}

Here, I present a series of laboratory activities that use the Model-Observe-Reflect-Explain Thinking Frame as an instructional tool to help facilitate students' thinking about the relationships between the evidence they collect in the laboratory and molecular-level behavior. Building on students' understandings of electrical conductivity (from a previous laboratory experience) as an indicator of dissolved ions in aqueous solution, we use conductivity as key evidence to inform students' models of the chemical systems they study.

2.2 Experimental Structure

2.2.1 Activities Preceding Precipitation Experiments

The laboratory experiment described in this chapter was designed to take place during the third week of a three-week laboratory module, following the laboratory experiments comprising the “What Happens When Chemical Compounds Are Added To Water?” module by Mattox, Reisner & Rickey,²⁹ which will be referred to henceforth as the “Dissolution” laboratory module. These experiments help develop students' understanding of the molecular-level behavior of solutions containing electrolytes and nonelectrolytes using conductivity as a key experimental evidence.

The Dissolution module centers on students' models of salt (NaCl) and sugar (C₆H₁₂O₆) both before and after they are added to water. In these experiments, students

observe the addition of number of substances to water – ionically-bonded compounds, covalently-bonded compounds, and pure metals – and use their observations of mixture appearance and measurements of conductivity to draw the inferences that substances of differing compositions and differing bonding types behave differently when added to water. Students are then asked to extend their reasoning to other situations, such as solutions with varying ionic concentration, or those containing polyatomic ions (NaNO_3) or nonmetal ions (NH_4Cl).

This module has been shown to be successful at prompting students to revise their models to be both more scientifically correct and more consistent with experimental evidence. In one analysis of students' models,¹² 35% began the laboratory module with a scientifically correct description of the dissolution of salt (NaCl) in their initial model, but 80% presented a scientifically correct view in their refined model.

Although it is desirable that students carrying out this laboratory module have completed the previously-published portions of the “What Happens When Chemical Compounds Are Added to Water?” module, this is not absolutely necessary; to interpret the evidence encountered in this experiment, it is most important that students possess a familiarity with the relationship between ion concentration and solution conductivity.

The new experiments on precipitation reactions that have been integrated with the previously-developed laboratory module build upon the foundation of these first two weeks, after which many students have arrived at the conclusion that the conductivity of a solution increases with an increase in the concentration of dissolved ions present. The third week of the “What Happens When Chemical Compounds Are Added to Water?” laboratory module extends this concept to students' models of reactions in aqueous solutions – specifically, precipitation reactions in which the concentration of dissolved

ions is reduced by the removal of ions from solution. The full texts of these activities are included as Appendix A, with the instructor's guide as Appendix B.

2.2.2 Initial Model Assignment

The initial model assignment asks students to extend their model from the previous weeks of the laboratory module to describe two reactions of aqueous solutions both before and after mixing. The first reaction is the mixing of lead nitrate and potassium iodide solutions, which produces a striking bright yellow precipitate of lead iodide. Although this reaction is often used in general chemistry textbooks as an illustration of precipitation reactions, students are typically not familiar with it and very few predict the macroscopic outcome of this reaction correctly. The second reaction students are asked to describe is the mixing of sodium hydroxide (NaOH) and potassium sulfate (K_2SO_4). This reaction does not result in a precipitate, however, students are also typically not aware of this based on their prior knowledge.

As implemented, this laboratory module has typically preceded instruction on precipitation reactions or solubility rules in the "lecture" component of the course. Students are encouraged not to refer to their textbook, solubility tables, or other references, as their initial models should reflect their personal understanding of what happens in the chemical systems.

Some common ideas exhibited by students in their initial models include: formation of precipitates that settle to the bottom of the reaction vessel (sometimes represented correctly, as ionic compounds, sometimes incorrectly, as "molecular" compounds); uncertainty about what products, if any, will be produced; or metathesis reactions involving the two dissolved reactants, leaving two "molecular" species

dissolved in solution. Instructors are encouraged to begin the laboratory period with a discussion of student models, asking several students with contrasting ideas to present their ideas for the class. At this point, it is important that instructors do not validate particular students' models as correct or incorrect, but they may ask students to discuss what experimental evidence would help support or refute particular ideas (particularly molecular-level) in the models. Aside from visual evidence, students who have just completed the "What Happens When Chemical Compounds Are Added to Water?" module often cite conductivity measurements as a potential source of information about molecular-level behavior.

2.2.3 Part I: Lead Iodide

The first set of experiments asks students to investigate three variations of the lead iodide precipitation reaction. Students mix 0.5 *M* solutions of lead nitrate with 0.5 *M* solutions of potassium iodide in differing volume ratios (2:1, 1:1, and 1:2), all resulting in the same total volume of 12.0 mL (Table 2.1). Solutions of the same molarities are used within a set of reactions to encourage students to focus on the ratios involved and the molecular-level behavior of the substances rather than on mathematical calculations involving concentration and quantity, consistent with recommendations to focus the systems studied in MORE laboratory experiments on simpler systems to avoid excessive cognitive demands that are likely to distract students from focusing and reflecting on molecular-level ideas and their relationships with the evidence.⁹

The different reaction mixtures produce differing quantities of precipitate (Figure 2.1) and – because of differing ion concentrations – different conductivities. Students

also measure the conductivities of the two reactant solutions prior to mixing. Ideal quantities of precipitate and ion concentrations are reported in Table 2.1, along with typical conductivity readings obtained in student laboratories using Vernier probeware.³⁰ Other models of conductivity meters yielding only semi-quantitative measurements have also been successfully used in this laboratory experiment.

Although the different quantities of precipitate formed are usually immediately apparent to students (Figure 2.1), they can be confirmed by centrifuging the reaction test tubes, and then quantified by removing the supernatant liquid with a transfer pipette and drying the solid samples to removed remaining water. Students also measure the conductivity of the supernatant liquid.

While samples are drying, the instructor leads a whole-class discussion in which student laboratory groups share their data with each other, and compare results and patterns. Students reliably observe that the lowest conductivity measurements are found in the mixture associated with the reaction that produced the greatest amount of precipitate – the stoichiometric reaction mixture. Non-stoichiometric mixtures, each containing leftover reactants, result in more ions being left in solution, yielding a higher conductivity.

The measurement of the masses of solids is not central to the experiment and has sometimes been omitted due to time constraints. However, measurement of the masses also allows students who have mastered stoichiometric calculations involving limiting reactants to reason about the identity of the precipitate based on its mass.

Table 2.1. Summary of typical observations and conductivity measurements for reactant and product mixtures.

Reaction	Reactant 1 quantity and conductivity ($\mu\text{S}/\text{cm}$)	Reactant 2 quantity and conductivity ($\mu\text{S}/\text{cm}$)	Identity and appearance of centrifuged precipitate	Conductivity of supernatant solution ($\mu\text{S}/\text{cm}$)
A	8.0 mL 0.5 M KI (~12400)	4.0 mL 0.5 M $\text{Pb}(\text{NO}_3)_2$ (~16900)	PbI_2 Smallest quantity of yellow precipitate	~13400
B	6.0 mL 0.5 M KI (~12400)	6.0 mL 0.5 M $\text{Pb}(\text{NO}_3)_2$ (~16900)	PbI_2 Moderate quantity of yellow precipitate	~11300
C	4.0 mL 0.5 M KI (~12400)	8.0 mL 0.5 M $\text{Pb}(\text{NO}_3)_2$ (~16900)	PbI_2 Greatest quantity of yellow precipitate	~9000
D	2.0 mL 0.02 M NiSO_4 (~2900)	4.0 mL 0.02 M NaOH (~ 5200)	$\text{Ni}(\text{OH})_2$ Small quantity of blue precipitate	~1900
E	3.0 mL 0.02 M NiSO_4 (~2900)	3.0 mL 0.02 M $\text{Ba}(\text{OH})_2$ (~9000)	$\text{Ni}(\text{OH})_2 + \text{BaSO}_4$ Greatest quantity of light blue precipitate	~10
F	2.0 mL 0.02 M K_2SO_4 (~5200)	4.0 mL 0.02 M NaOH (~ 5200)	No precipitate	~5200
G	3.0 mL 0.02 M K_2SO_4 (~5200)	3.0 mL 0.02 M $\text{Ba}(\text{OH})_2$ (~9000)	BaSO_4 Smallest quantity of white precipitate	~5100

Reflection questions embedded throughout the laboratory experiment ask students to offer a molecular-level explanation for their explanations, posing questions such as “Based on your data and observations, what happens *on the molecular level* when $\text{Pb}(\text{NO}_3)_2$ (aq) and KI (aq) are mixed? Discuss this with your group.” and “Are your experimental results consistent with your initial model? If your results are not consistent with your model, how would you revise your model?” Students are therefore prompted to revise their model to be consistent with their evidence, supporting

molecular-level models in which some ions come together to form a visible and quantifiable solid precipitate, while other ions remain in solution.

2.2.4 Part II: Sulfates and Hydroxides

The second portion of the laboratory experiment involves a system of four reactions, summarized in Table 2.1. Reaction F produces no precipitate; reactions D and G produce one precipitate each, while reaction E produces two precipitates – nickel hydroxide and barium sulfate.

In this portion, students only perform one trial of each reaction, but must first write a balanced chemical equation in order to determine the stoichiometric ratio and therefore the volumes of each solution needed in order to produce equal total volumes of each reaction mixture. Students therefore identify chemical formulas of potential products, a task that helps them think about the molecular-level behavior of the ions that make up these compounds as they observe the results of each reaction. As with the first set of reactions, students also record the conductivity of reactant and product mixtures, as well as observe and centrifuge the resulting mixtures.

As in Part I, centrifuging the reaction mixtures allows students to clearly compare both the colors and the quantities of precipitates formed. Figure 2.1 provides pictures of the four reaction mixtures after centrifuging. The greatest quantity of precipitate is formed in reaction E, where – if the correct stoichiometric ratio is used – essentially all ions should precipitate out of solution into one of the two solids. Reaction F produces no precipitate, which leads students to conclude (without reference to solubility rules) that both potassium hydroxide (KOH) and sodium sulfate (Na_2SO_4) are soluble. By process of elimination using the balanced chemical equations that students have written,

students can conclude that $\text{Ni}(\text{OH})_2$ precipitates in reaction D, and BaSO_4 precipitates in reaction G.

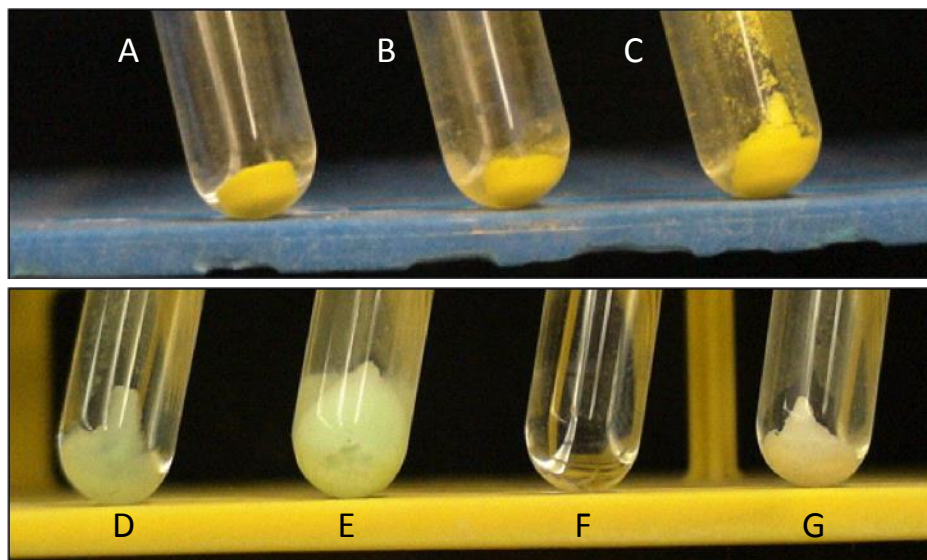


Figure 2.1. Images of each product mixture after centrifuging. Top: Product mixtures A, B, and C from part I, illustrating different stoichiometries resulting in different quantities of lead iodide precipitate. Bottom: Product mixtures from part II, containing: (D) nickel(II) hydroxide; (E) nickel(II) hydroxide and barium sulfate; (F) no precipitate; (G) barium sulfate.

Although we have found that most students have not encountered a system in which multiple products of a reaction precipitate, several pieces of evidence point toward that unexpected conclusion regarding this reaction. First, the quantity of precipitate in reaction E is greater than in the others; additionally, the color lies between the white of BaSO_4 and the green of $\text{Ni}(\text{OH})_2$. Both of these observations suggest that the precipitate is a mixture of both substances. Finally, the very low conductivity of the supernatant solution offers additional evidence that practically all of the ions have precipitated out of solution.

Again, a class discussion is typically conducted to summarize the data. Because differing ion concentrations and ionic mobilities result in differences in conductivities of reactant solutions with comparable ionic concentrations – a topic not directly addressed in this experiment – we have sometimes encouraged students to predict what the mixture conductivity would be if calculated as an average of the conductivities of the two original solutions. Comparing the measured conductivities with these predicted values provides students with additional evidence regarding the nature of the reaction that has taken place.

2.2.5 Part III: Student-Designed Experiments

A final experiment asks students to conduct variations on reaction E – the mixture producing two precipitates – using different ratios of reactants. This experiment is less explicitly guided and does not specify reaction ratios, but allows students to further explore these systems and obtain further evidence to inform their models.

As students test ratios that diverge from the stoichiometric ratio, they observe a smaller quantity of precipitate formed and, correspondingly, higher solution conductivity. These data can serve to emulate the data generated from a conductimetric titration curve.^{23b, 31} We have found that, due to time constraints, students rarely have time to conduct such experiments during a standard three-hour lab period, but they are sometimes of interest to fast-working groups or highly-motivated students. As an abbreviated exercise, some instructors have asked students to predict (without performing the experiment) what patterns they would expect in both precipitate quantity and solution conductivity for different reactant ratios.

In the laboratory curriculum at Colorado State University, this laboratory module, because it touches on the effects of reaction ratios, has been used prior to the “Exploring Gold Nanoparticles” laboratory module,³² where the initial model assignment asks students to predict the outcome of a nanoparticle synthesis reaction in which different ratios of reactants are used.

2.3 Outcomes: Refined Models

Does participation in these laboratory activities result in students producing refined molecular-level models that are more scientifically accurate than their initial models?

To determine whether these laboratory activities fostered an improved understanding of the molecular-level behavior of these systems, one laboratory section during the fall of 2007 – taught by the author – was selected for analysis. This laboratory section consisted of chemistry majors, 20 of whom consented to have their laboratory reports analyzed.

Students’ models – both of the two reactant solutions and of the product mixtures – were coded using a scheme previously developed¹² for characterization of students’ ideas about salt and sugar solutions in the Dissolution laboratory module. To validate its use by this researcher, 17 models from a previously-coded student population (students at a research university) were analyzed. This duplicate coding obtained an intercoder reliability (Cohen’s kappa) of 0.79, which can be described as “substantial” agreement.³³ A summary of the results of the coding is outlined in Table 2.2. A greater proportion of students exhibit scientifically correct ideas about both

chemical systems discussed in their refined models than in their initial models, although the differences are not statistically significant in this sample ($p > 0.100$ for both systems).

Table 2.2. Student conceptions about the products of the precipitation reactions in initial and refined models surrounding the laboratory experiments described in this chapter. Scientifically correct descriptions of the reactions in the refined model are indicated with “*”.

	Pb(NO₃)₂ + KI	NaOH + K₂SO₄
Initial models:		
Correct conceptions	25%	35%
Refined models:		
One product precipitates	55%*	0%
Incorrect description with precipitate	5%	10%
Ions in solution	5%	60%*
Molecules in solution	20%	5%
Other misconceptions/vague	15%	25%

Figure 2.2 presents some examples of the illustrations produced by a student (07-795) who refined her molecular-level model to be consistent with experimental evidence. In the top picture, taken from her initial model, the student illustrates separated ions existing “during” the reaction (immediately following the mixing of solutions), she depicts both products as molecular species in solution. However, in her refined model (the bottom picture in Figure 2.2), she represents potassium nitrate as remaining as separated ions while lead iodide forms a precipitate that settles on the bottom of the reaction vessel.

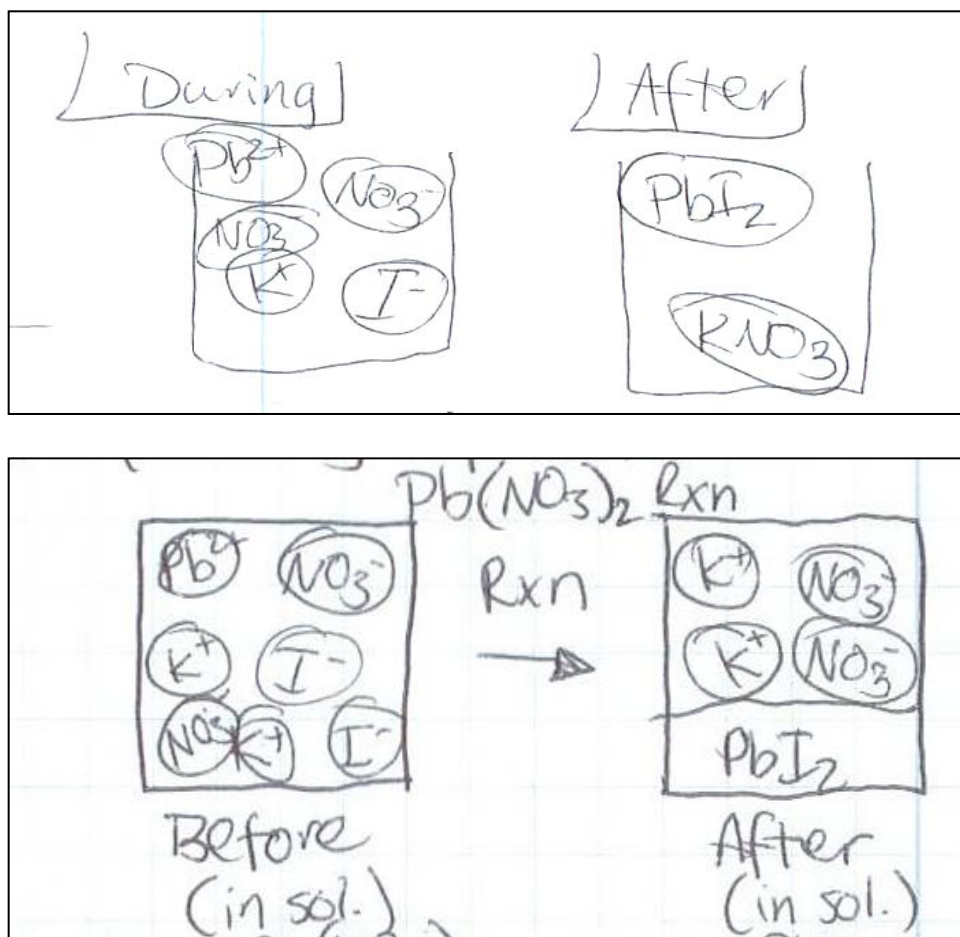


Figure 2.2. Molecular-level pictures from student 07-795’s initial (top) and refined (bottom) models of the lead nitrate/potassium iodide reaction, illustrating – particularly for the product mixtures labeled in both as “after” – refinement of her personal model towards greater scientific correctness.

2.4 Effects of Context

2.4.1 Conductivity Context

The work described in this section is also described in the paper: Teichert, Melonie A.; Tien, Lydia T.; Anthony, Seth; Rickey, Dawn, Effects of context on students’ molecular-level ideas. *International Journal of Science Education* **2008**, 30 (8), 1095-1114. The primary contribution of this dissertation’s author to that paper, as described

here, was the design, implementation, and analysis of the “conductivity prompt” experiment described below.

During previous studies of the “What Happens When Chemical Compounds Are Added to Water?” laboratory module, and subsequent student interviews, an unusual phenomenon was observed in which students expressed scientifically correct ideas about a chemical system in one context, and yet, only about fifteen minutes later, would express a contradictory, incorrect idea about that same chemical system within a different context.

In this work, described in detail in Teichert, Tien, Anthony & Rickey,³⁴ students enrolled in a general chemistry course at a community college and at a research university both completed the first two weeks of the Dissolution module, and, near the end of the academic term, were interviewed. Among the components of the interview was a question in which students were asked to predict the conductivity of various solutions, including $\text{NaCl}(aq)$, and asked to draw a molecular-level picture of those solutions. Later within the same interview, students were given a question in a new context that they had not encountered within their chemistry course – colligative properties. After reading a paragraph describing how boiling point elevation is a function of the number (but not the type) of dissolved particles, the interviewed students were asked to predict the relative boiling points of several pairs of solutions, including $0.02\text{ M C}_6\text{H}_{12}\text{O}_6$ and 0.02 M NaCl , and were also asked to draw molecular-level pictures of these solutions. Of the 19 students interviewed, 95% described $\text{NaCl}(aq)$ as existing as separated ions in the context of conductivity; however, later, when asked about an $\text{NaCl}(aq)$ solution in the context of boiling point elevation, only 53% did so, a significantly smaller proportion ($p=.008$, two-tailed Fisher’s exact test).³⁴

Several interpretations of this data are possible. Students may have been cued to activate ideas relating to the laboratory because the context of the interview question evoked associations with the laboratory itself. However, characteristics of the context itself may also be at play – the idea of conductivity could activate different student conceptions. To investigate these hypotheses, the author of this dissertation, upon first learning about the above results, proposed and executed an experiment to investigate the effect of the context itself prior to students' exposure to the laboratory experiment.

To this end, a population of students drawn from chemistry honors and majors' sections at a research university were assigned one of two different versions of the initial model assignment within the Dissolution module. While one group ($n=28$) received the standard model assignment, another group ($n=14$) received the model assignment with "conductivity prompts" included. In the model assignment that included these prompts, students were asked not only to describe their macroscopic sensory and molecular-level expectations about salt and sugar added to water, but also to make predictions about the conductivity measurements that they would take, and to explain those predictions with reference to the molecular level.

The initial model assignment²⁹ is included here; the "conductivity prompts" present in only some students' models are underlined:

Describe your understanding of what happens to chemical compounds when they are added to water. For your initial model, discuss salt (NaCl) and sugar (C₁₂H₂₂O₁₁), two solids that can be found in any kitchen cupboard. Describe what you expect to observe (see, hear, feel, smell) before and after you have added salt to water AND before and after you have added sugar to water. In addition, include your expectations about

the measurements that you will take on these solutions with the conductivity tester. This is your *initial macroscopic model*. Then, explain what you think the molecules, atoms, and/ or ions are doing that results in your observations; this is your *initial molecular-level model*. When developing your model of what happens when chemical compounds are added to water, be sure to incorporate responses to the following questions/statements into your model. (Your instructor will be looking for these.)

- What kinds of changes do you expect to observe when you put each substance in water?
- How will the amount of substance added affect your observations?
- Using the language of chemistry, explain what is happening to your compounds as they are added to water.
- From a molecular-level perspective, explain your predictions about conductivity.
- Draw pictures of what you expect the atoms, ions, and/or molecules to look like in solution. (Molecules, atoms, and ions can be represented with symbols. If you use symbols, be sure to use a key.)

The researcher analyzed student models using the same coding scheme used in previous work¹² to classify student conceptions; the results are summarized in Table 2.3.

Table 2.3. Frequent student descriptions of sodium chloride within initial models, for the “What Happens When Chemical Compounds Are Added to Water?” in response to initial model assignments both with and without a “conductivity prompt.”

	NaCl as “molecules”	NaCl as “ions”	Other conceptions
Model assignment without conductivity prompt (n=28)	43%	25%	32%
Model assignment with conductivity prompt (n=14)	7%	65%	28%

While 65% of the students who received the conductivity prompts as part of their model assignment represented NaCl(aq) in their written initial models as separated ions, only 25% of those who did not receive the prompts did so – a statistically significant difference (two-tailed Fisher’s exact test, $p=0.020$). Additionally, a greater fraction of student models without the conductivity prompt represented NaCl(aq) as neutral molecules (two-tailed Fisher’s exact test, $p=0.032$).

These results reflect a similar phenomenon to that seen in the interviews – the context in which students’ ideas are solicited can have a profound impact on what ideas they express. Moreover, since this is seen prior to students conducting the laboratory experiment, it indicates that the nature of the context itself, rather than a particular instructional experience, may be a primary factor. The context of electrical conductivity may activate certain ideas about charge and charged particles that might not be activated in a different context, such as that of boiling point elevation – specifically, ideas about electrolytes separating into their constituent ions in solution. Independent of instruction, the nature of the context itself appears to be important.

2.4.2 Precipitation Reaction Context

Because the laboratory experiment on precipitation reactions described in this chapter was coupled with the “What Happens When Chemical Compounds Are Added to Water?” module, the final refined model assignment for that module and the initial model assignment for this experiment were conjoined. After – after presenting their refined models of salt and sugar solutions, with description of changes, students were then asked to present their model of the two precipitation reactions discussed previously – the mixing of lead nitrate solution with potassium iodide solution, and the mixing of sodium hydroxide solution with potassium sulfate solution.

This provided an opportunity to examine students’ conceptions of ionic compounds dissolved in water in two parts of the same model – students’ descriptions of salt solutions in their refined models and students’ descriptions of the reactants in their initial model of precipitation reactions. In these two juxtaposed situations, we also observed context-dependence of students’ ideas.

Student models from multiple sections taught by graduate teaching assistants at Colorado State University in the fall of 2007 were collected. Models from 105 consenting students were analyzed; of these, it was found that many students omitted the “expand your model” section of the assignment, or omitted molecular-level descriptions of reactants from their models. For the 64 students who presented a molecular-level model for reactant solutions, we analyzed their descriptions of the reactant solutions for the lead nitrate and potassium iodide reaction.

Of the 64 students, 19% ($n=12$) described all reactants as intact “molecular” species in solution and 52% ($n=33$) described all reactants as separated ions in solution.

29% ($n=19$) articulated other conceptions, or did not describe both reactant solutions similarly.

An example of “molecules” conceptions from a student model is illustrated in Figure 2.3. Similar misconceptions have been reported by Kelly et al.^{15b}

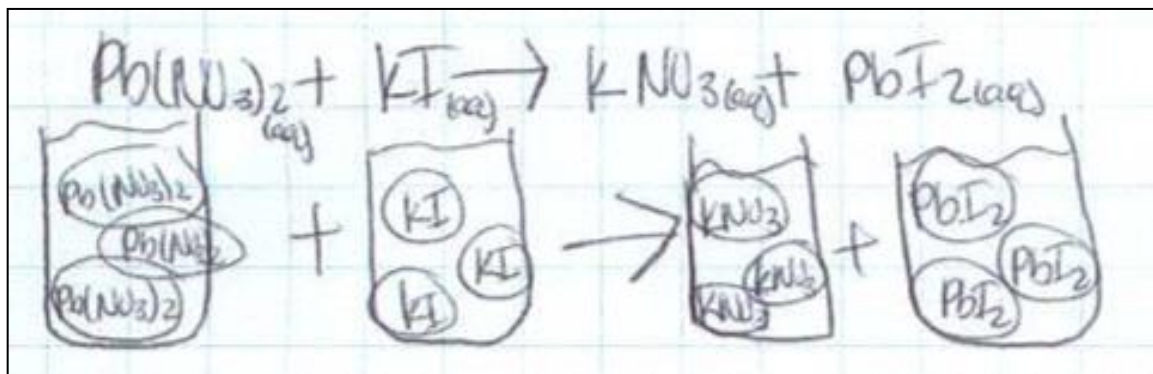


Figure 2.3. Molecular-level pictures from Student 07-253’s initial model of the lead nitrate/potassium iodide reaction, illustrating a “molecules” conception. The student depicts both reactants and products as existing as dissolved neutral molecules.

If students possessed coherent and consistent underlying models of ionic compounds breaking apart into ions in solution, or as “molecules” in solution, we would expect students to exhibit similar ideas in both their models of sodium chloride (as part of the Dissolution final model) and of lead nitrate and potassium iodide (as part of the precipitation initial model). However, their conceptions do not appear to be stable across this change in contexts – as seen in previous analysis of the Dissolution model, the overwhelming majority of students express a scientifically correct “ions” conception in their refined models. This contrasts with the “molecules” conception they express in this initial model assignment that immediately follows it. Indeed, analysis of the student ideas contained within these models reveals that, of the 33

students who expressed an ions conception in their descriptions of reactant solutions, 85% ($n=18$) previously described sodium chloride as dissolved ions. However, a comparable percentage, 75% ($n=9$) of the 12 students who described reactant solutions as dissolved molecules, did so as well. These students, similar to many of the interviewed students, expressed contradictory views of the dissolving of ionic compounds in different contexts.

Models of the product mixtures for both reactions were also examined, although the ideas expressed there were more varied. Of the set of students who included molecular-level ideas, 20% expected the product mixtures to contain only dissolved ions. Of these, 95% described reactant solutions as dissolved ions. Students therefore described similar molecular-level conceptions of reactant and product mixtures in the context of these reactions. Figure 2.4 illustrates this with an example from a student model that predicts that both reactant and product mixtures will contain dissolved ions.

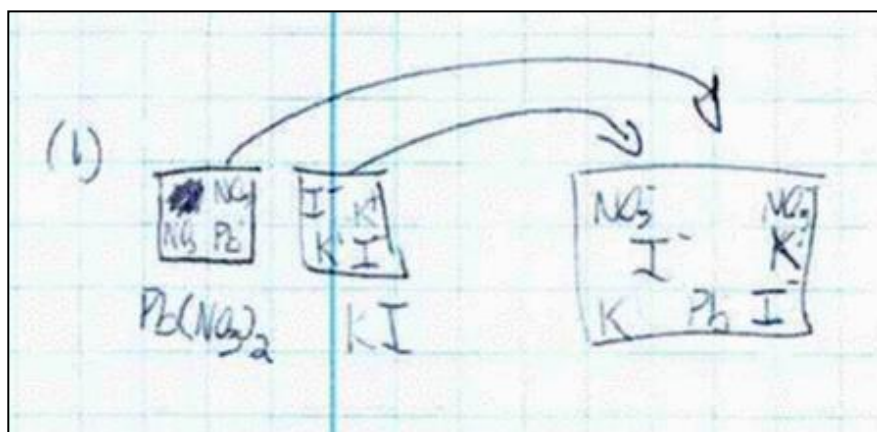


Figure 2.4. Molecular-level pictures from student 07-223's initial model of the lead nitrate/potassium iodide reaction, illustrating an "ions" conception. The student depicts both reactant and product mixtures as containing dissolved ions.

Similarly, of the 17% of students whose models of product solutions described products as dissolved molecules (with products KNO_3 and PbI_2 predicted by metathesis of reactants), 83% of these students had described the reactant solutions as containing molecules. The example shown in Figure 2.3 illustrates these student conceptions.

2.5 Discussion

In summary, although differences in context can result in students activating different molecular-level conceptions and therefore expressing conflicting or contradictory ideas, there is also evidence that students activate similar ideas within similar contexts, creating models of both reactant and product mixtures that exhibit similar features.

As in the interviews, a substantial number of students expressed one idea – a scientifically correct one, consistent with the evidence they encountered in the Dissolution module, in the refined model portion of this assignment, but then failed to apply that same model when asked to describe solutions of ionic compounds as reactants in the context of a precipitation reaction later in the same assignment.

Results such as these indicate that, while students may activate ideas consistently within certain contexts, when those contexts are changed, they may not activate those ideas, or they may activate different ideas. Similar phenomena have been observed in the field of physics and cited as evidence to support a theory of the structure of knowledge in the mind that is sometimes called “knowledge in pieces,”³⁵ in contrast with a competing perspective³⁶ that students’ ideas about scientific concepts are, at their core, conceptually coherent frameworks.³⁷ Fragmented “knowledge in pieces” has been

offered as a partial explanation for students' "shallow" learning and failure to transfer their knowledge into new contexts.

Students' sensitivity to context illustrates one aspect of the "shallowness" of many students' understanding that instructional approaches such as the MORE Thinking Frame have been found to address. Analysis of the interviews described above, in conjunction with students' models, though, revealed important relationships between three characteristics of students' models and their success at transferring their understanding into new contexts, as illustrated by success at the boiling point elevation question. These three characteristics are:¹³

- presenting a refined molecular-level model that is consistent with experimental evidence;
- correctly describing whether and how one's molecular-level ideas have changed from the initial model (this description of how one's own ideas have changed is a facet of metacognition); and
- citing specific experimental evidence to justify either refining one's molecular-level model or keeping it the same as part of the metacognitive reflection.

However, in this study, only 32% of students explicitly engaged in all three of these thinking processes in their refined model. On one level, this relatively low level of "success" is quite understandable – for many of these students, it was the first time they had been asked to write such a model, and the thinking processes involved in doing so are challenging. On the other hand, the difficulties that students face in engaging in writing these models motivated further work at developing activities to help students better understand what was expected of them in these model-writing assignments. One of the outcomes of this work was the development and assessment of a new set of

introductory activities, implemented at the start of the semester in a laboratory course using the MORE Thinking Frame, and described in the following chapter.

Chapter 3: Investigating the Use of Modeling and Critiquing Activities to Enhance Student Engagement in Scientific Thinking Processes

The thinking processes involved in constructing and revising explanations based on experimental evidence lie at the heart of scientific practice and have also been identified by multiple professional scientific organizations as key learning goals.³⁸ A growing number of inquiry-based chemistry curricula and instructional approaches emphasize the importance of developing these abilities.^{5b, 5c, 39} Inquiry-based laboratory approaches – such as Model-Observe-Reflect-Explain (MORE)⁹⁻¹⁰ and the Science Writing Heuristic⁴⁰ – encourage students to develop scientific explanations based on evidence, and have been shown to improve student understanding of chemistry concepts when compared with more traditional, verification approaches.^{9-10, 41}

While the benefits of inquiry-based learning can be substantial, students often struggle to engage in thinking processes that are integral parts of these activities. In a range of disciplines, when college students are asked to construct scientific explanations, they frequently fail to support claims with sufficient evidence, explain the relationship between evidence and claims,⁴² or show awareness that prior knowledge plays a role in the interpretation of evidence.⁴³ In addition, when asked to express ideas about chemical systems, students often fail to refer appropriately to the particulate level of atoms, molecules, and ions, the basic explanatory units of chemistry.⁴⁴

Chemistry educators have suggested strategies that may help students engage in key thinking processes involved in inquiry,⁴⁵ and designers of inquiry-based college

chemistry laboratory instruction have devoted special effort to confronting these challenges by explicitly defining and explaining to students what constitutes a valid scientific argument or by providing students with scaffolding as they engage in their first attempts to construct such explanations.^{29, 46} Other efforts in a range of contexts and educational levels have had students critique high-quality examples or peers' efforts.⁴⁷ Use of peer-critiquing systems such as Calibrated Peer Review has been shown to improve the quality of students' writing in science courses.⁴⁸ Structured critiquing activities therefore have promise to help students understand the cognitively-complex tasks associated with inquiry. It is important, however, to assess the effectiveness of particular approaches in the context of inquiry-based instruction.

In this chapter, we present the *Learning to Construct and Refine Models* activities, which were designed to teach students to use the MORE (Model-Observe-Reflect-Explain) Thinking Frame to develop and refine personal models of chemical systems, as well as results of assessments of their effectiveness. These activities have undergone multiple refinements, and in each implementation, students produced written models describing their ideas about a burning candle before and after the wick is lit. To assess the effectiveness of the activities and the impact of year-to-year revisions – including the addition of model-critiquing activities – we analyzed students' written models for evidence of their engagement in several key scientific thinking processes related to the key aspects of cognition identified in previous research¹³ and described in Chapter 1.

3.1 *Learning to Construct and Refine Models Activities*

The first implementation of *Learning to Construct and Refine Models* in 2006 consisted of a one-hour introduction to the MORE Thinking Frame during the first laboratory period of the semester. In 2008, it was expanded to a full laboratory period of two hours and fifty minutes. Below, we describe the 2006 implementation and then outline revisions made through 2008. The full text of the most recent implementation is included as Appendix C, with the text of the instructor's guide for this activity in Appendix D.

2006 implementation. During each implementation of *Learning to Construct and Refine Models*, instructors began with a brief verbal explanation of the MORE Thinking Frame, outlining its four major components (Model, Observe, Reflect, Explain) and the expectations for the written model assignments. Students then participated in a series of activities leading them through a cycle of applying the MORE Thinking Frame: students wrote an initial model, discussed and critiqued peers' initial models, conducted an experiment, and wrote a refined model.

The initial model assignment asked students to describe a familiar system – a candle before and after the wick is lit – from both macroscopic and molecular-level perspectives. Figure 3.1 contains an abbreviated version of this assignment. This system was chosen because, while being within the realm of most students' experience, it offers a variety of phenomena for consideration at both the macroscopic and molecular levels. After writing their models, students exchanged models within peer groups, identified major aspects of each model (presentation of macroscopic and molecular-level ideas, discussion of changes, use of supporting evidence), ranked their group member's

models in terms of degree of completeness of each aspect, and suggested improvements to the original authors of each model.

Next, students conducted experiments to obtain evidence that could inform their models. First, they observed a candle both before and after it was lit, as discussed in their initial models. Students then surrounded the lit candle completely with an inverted flask or beaker while the candle sat in a pan of water. Once the lip of the flask was below the surface of the water, students observed the flame extinguish and the water level in the flask rise. In this second experiment, students obtained evidence that many of them had not considered when writing their initial models. [Note: A common misconception regarding this experiment, addressed by later experiments in the laboratory course, is that the water level within the flask rises due to the consumption of oxygen. Careful observation reveals that the majority of water level rise follows the extinction of the flame, and is due primarily to the cooling of the hot air trapped in the flask.⁴⁹]

After completing their experiments, students wrote refined models. The refined model assignment, provided in abbreviated form in Figure 3.1, asked students to again describe their understanding of the system from both a macroscopic and molecular-level perspective and to discuss what changes they made to each key aspect of their model. For their molecular-level models, students were asked to describe how experimental evidence supported those changes – or, if no changes were made, how their experimental observations supported their initial model. As time permitted, students shared their refined models with their peers, who were instructed to leave written feedback for the model writer.

Model-Writing Assignments

Initial Model Assignment:

Describe your understanding of what happens when a candle burns. Describe what you expect to observe (see, hear, feel, smell) before and after the candle wick has been lit; this is your *initial macroscopic model*. Then, explain what you think the molecules, atoms, and/or ions are doing that results in your observations; this is your *initial molecular-level model*. Describe how you think your molecular-level model is connected to or explains your macroscopic model. In addition to words, you may choose to use pictures to represent your models.

Refined Model Assignment:

Develop a refined model for what happens when a candle burns. For a candle before and after the wick is lit, briefly describe your refined macroscopic model (your observations) and your refined molecular-level model that accounts for your observations. Compare your refined model with your initial model, and identify the key aspects of your model that changed or remained the same. Be specific. Then explain why your model has changed in these ways (or why it has not, if aspects of your refined model are exactly the same as in your initial model) by using specific experimental evidence to support or refute the claims you made in your initial model.

Figure 3.1. Abbreviated assignment instructions, as implemented in 2008 during the *Learning to Construct and Refine Model* activity. The full assignment is included within complete activity in Appendix C.

2008 revisions. In the 2008 implementation of *Learning to Construct and Refine Models*, additional activities, focused on students' reviewing and critiquing of sample models, were integrated following the introduction to the MORE Thinking Frame. Students read a typical MORE initial model assignment and several sample initial models provided in their laboratory manuals. (See Appendix C). These initial models, and the sample refined models reviewed later by the students, addressed a simple chemical system – the mixing of equal volumes of ethanol and water. The sample models exemplified many of the characteristics desired in student models. They explicitly identified molecular-level particles and described their behaviors, connecting macroscopic predictions and molecular-level behavior. However, to create contrasts to serve as catalysts for discussion, and to directly confront commonly-observed student difficulties, these models also contained several intentionally-introduced deficiencies, such as ambiguities in molecular-level language or omissions of key aspects of models.

As students read the sample models, they evaluated them using rubrics provided in their laboratory manuals. These rubrics, which paralleled those that were used by instructors to grade student work, required students to rate each aspect of the model as “not present,” “present but incomplete,” or “present and complete.”

After working individually, students were encouraged to compare their assessments of the sample models in small groups and attempt to reach consensus on their judgments. The instructor then led a whole-class discussion, soliciting student analysis of model quality, prompting students to discuss judgments on which they disagreed, and providing feedback on students' evaluations.

After discussing sample initial models, students observed the instructor perform the experiment that the models described – the mixing of equal volumes of ethanol and

water. Students observed that the mixture had a total volume less than the sum of the volumes of the two components before mixing, which was unexpected for many students. Students then reviewed sample refined models (also included in Appendix C). Like the sample initial models, these refined models exhibited a range of quality and included some intentional deficiencies, including lack of clarity about changes made to the model and ambiguities in referring to supporting experimental evidence. The new aspects of the *Learning to Construct and Refine Models* activity that were added in 2008 – reviewing, critiquing, and discussing sample models – lasted approximately an hour and fifteen minutes.

The 2008 implementation also incorporated greater scaffolding for students' reviews of peers' models compared with the 2006 implementation. When students exchanged their written models with peers, they were provided with a rubric similar to that used in reviewing sample models. Students then provided brief written feedback to their peers and discussed their ratings with the models' original authors. (In some laboratory sections, peer review of refined models was curtailed due to time constraints.) During the revised activity, students also completed a post-laboratory assignment to further refine their models, but these models were not analyzed, since the 2006 cohort did not complete this assignment.

3.2 Participants and Methods

During each year of implementation of *Learning to Construct and Refine Models*, one or more sections of a first-semester general chemistry laboratory course taught at the same large research university in the United States were selected for study.

In the fall of 2006, the study population was drawn from two sections of chemistry majors and one section of honors students (the only sections participating in MORE laboratories), and in the fall of 2008, from four sections of regular science and engineering students (with majors other than chemistry). The same two graduate teaching assistants (GTAs) taught these sections during both semesters, and both of the GTAs were involved in research on laboratory instruction employing the MORE Thinking Frame. In the following year, 2009, a laboratory section taught by one of these GTAs was also studied.

The study included only students who consented to participate and who completed pre-course demographic surveys. Additionally, because not all students completed the refined candle model during the laboratory session, we excluded the models of students who did not complete the portions of the models that were analyzed. The resulting cohort size for each year are reported in Table 3.1, as are the percentages of students from the selected laboratory sections whose models were complete, which was comparable for the two populations.

During the first week of each semester, all students included in the study completed the General Mathematics and General Chemical Knowledge sections of the American Chemical Society's Examinations Institute 1998 Toledo Chemistry Placement Examination. Table 3.1 reports mean scores for each cohort. The 2008 cohort's mean score on the General Chemical Knowledge section of the exam was significantly lower than that of the 2006 cohort of chemistry majors and honors students ($t(32)=2.37$, $p=0.012$), consistent with the groups' academic backgrounds and levels of interest in chemistry. There was no statistically-significant difference between the mean scores of the groups on the General Mathematics section of the exam.

Three key aspects of each student's refined candle model – quality of molecular-level language, quality of reflection on changes, and quality of use of supporting evidence – were coded independently by two researchers. Disagreements in coding between the two researchers were resolved through discussion to obtain the final codes employed in the analyses presented here.

3.3 Molecular-Level Language

Describing systems in terms of the behavior of molecular-level entities is a central aim of explanations in chemistry. In the context of MORE refined models, such descriptions are also prerequisite to describing how one's molecular-level model has changed and how evidence supports those changes. However, the language students used to describe chemical systems in their models varied widely, such that some could not, strictly speaking, be considered molecular-level models.

We characterized each student's refined candle model based on the most sophisticated instance of molecular-level language present in the model. In this coding scheme, the most sophisticated type of reference to molecular-level entities described “clear molecular-level entities and behaviors” – references to the behavior or properties of molecular-level entities consistent with a particulate view of matter. These descriptions contained references to identifiable particles as molecular-level entities (e.g. “oxygen molecules,” “carbon atoms”) and also contained descriptions of physical arrangement, position, or motion applicable to submicroscopic, particulate forms of matter (e.g., “atoms form a tightly packed lattice,” “molecules are close together,” “atoms move faster”).

Other students' descriptions, however, referred to the use, creation, or destruction of molecular-level particles in terms that could as readily apply to non-particulate or bulk matter ("is consumed," "is released," "reacts," "interacts"); or suggested that individual particles engage in macroscopic behavior ("burns," "warms up"). Descriptions that identified molecular-level particles in a manner not clearly consistent with a particulate conception were classified as referring to "clear molecular-level entities, macroscopic behaviors." Student models that did not contain descriptions that fit either of these categories were coded as having "no molecular-level reference."

3.4 Reflection on Model Revisions and Use of Evidence

We also coded students' refined candle models for the quality of two important characteristics of their reflection on refinements to their models that are strongly correlated at their success at applying their molecular-level models in new contexts:¹³ (1) accurate and complete reflection on whether and how their molecular-level ideas changed from the initial model to the final model, and (2) use of experimental evidence to justify their model refinements (or to support aspects of their models that did not change) as part of the reflection.

Refined Model Excerpts

Student 1:

“I was pretty accurate on my after model, however, I didn’t realize the oxygen molecules in the air were a big part of the reaction that created the flame and gave off heat. This was shown in the experiment when we sealed off the candle from the air which meant it only had a limited supply of oxygen molecules to react with. So when there wasn’t any oxygen left, the molecules in the wick stopped reacting with the oxygen molecules and the flame went out.”

Student 2:

“Changes to micro: There were no changes to the before part because my initial model was correct. There were changes to the after part because the information gained from the experiment needed to be included.”

Figure 3.2. Excerpts of contrasting quality from two students' refined models discussing changes to the student’s molecular-level model.

In coding students’ reflections on their model revisions, we classified them as “accurate” if the student specifically and correctly identified which aspects of their models changed; “somewhat inaccurate” if the student’s descriptions were specific, but contained inaccurate statements about how their model had changed; and “incomplete or vague” if the student alluded to changes, but was not specific about how their models

changed. Student 1's discussion of changes excerpted in Figure 3.2 was coded as “accurate,” as she correctly identified that her refined model included a discussion of the role of oxygen molecules in the combustion reaction that was not included in her initial model. Student 2's discussion of changes was judged to be “incomplete”. Although she acknowledged adding ideas to her molecular-level model, she failed to specify what new information about the molecular-level was included.

The refined model prompt also asked students to describe how experimental evidence supported refining their molecular-level models or keeping aspects of those models unchanged. Students' use of evidence to support these statements about their molecular-level models was classified as “specific” if students identified experimental evidence or observations that supported either changing or keeping aspects of their molecular-level model, and “vague” if students only alluded to some role of experimental evidence in model refinements without noting specific evidence. For instance, because Student 1 identified specific observations – the extinction of the candle flame – and connected them to changes to her model, her use of evidence was coded as “specific.” Student 2's discussion of evidence, however, was judged to be “vague” – although she indicated an awareness of the role of experimental evidence in prompting model revisions, she failed to cite any specific evidence.

3.5 Outcomes: Model Quality for 2006 and 2008 Cohorts

We anticipated that having students read, critique, and discuss example models, as well as providing additional structure for these and peer reviews in 2008 compared with 2006, would result in improvement in the quality of students' models along the three dimensions discussed above. Table 3.1 summarizes these data, and reveals a

reliable difference in quality of molecular-level language used in the models written by the two cohorts. (Mann-Whitney $U(32)=79.0$, $p=0.013$; the Mann-Whitney test compares the distributions of data within sets where data is ordered, but cannot be assumed to be parametric). Between the 2006 and 2008 implementation, improvements were also seen in quality of students' reflections on changes to their molecular-level models ($U(32)=89.50$, $p=0.059$) and on use of evidence ($U(32)=112.00$, $p=0.080$). These later differences are not as pronounced as the improvement in the use of molecular-level language, amounting to what is typically described as a "statistical trend." However, they are notable in light of the 2008's cohort's lower mean scores on the Toledo exam. The 2008 cohort displayed a higher level of proficiency despite a lower level of background knowledge in chemistry than 2006's population of honors students and chemistry majors.

Moreover, while no students in 2006 demonstrated fully specific and accurate reflection on changes to his/her molecular-level model or any use of supporting evidence, some students in the 2008 sample demonstrated both of these thinking processes in their refined models – an improvement that demonstrates the greater success of the 2008 version of the introductory activities at prompting students to engage in key thinking processes.

Table 3.1. Placement exam scores and quality of students' refined models along three key attributes for each cohort and model. Comparisons for each attribute with the 2008 *Learning to Construct and Refine Models* activity: * Statistically significant differences ($p < 0.050$)

† Statistical trends ($p < 0.100$).

	2006, Learning to Construct and Refine Models	2008, Learning to Construct and Refine Models	2008, Spectros- copy (first lab module)	2009, Learning to Construct and Refine Models
Population:	honors students & chemistry majors	regular science and engineering majors		
Fraction of original population with complete models	33%	30%		78%
Final study sample size	N=14	N=20		N=18
ACS Toledo Placement Exam:				
General Mathematics section mean	16.9/20	15.8/20		15.4/20
General Chemical Knowledge section mean	13.5/20	10.5/20		10.1/20
Use of molecular-level language:				
Clear molecular actors and actions	42%*	75%	90%	72%*
Clear molecular actors, macroscopic actions	0%*	20%	10%	17%*
No molecular-level references	57%*	5%	0%	11%*
Description of changes to molecular-level model:				
Accurate reflection on model changes	0%†	25%	55%*	33%*
Somewhat inaccurate reflection on changes	13%†	10%	15%*	11%*
Vague or incomplete reflection on changes	20%†	30%	20%*	44%*
No mention of changes or initial model	67%†	35 %	10%*	11%*
Use of evidence to support changes to molecular-level model:				
Cites specific evidence	0%†	15%	20%	12%
Vague reference to evidence supporting changes	0%†	5%	10%	6%
No use of evidence to support changes	100%†	80%	70%	83%

Although it would be desirable to evaluate the impacts of differing introductory activities on student performance on subsequent assignments, direct comparisons are not possible for the 2006 and 2008 cohorts because the schedule of later laboratory experiments differed. However, for the 2008 cohort, one researcher coded the quality of the next refined model produced by students in the laboratory course. In this assignment, students were asked to describe what happens in a hydrogen discharge tube after electrons are excited to higher energy levels.

Summary data for the 20 students from the 2008 cohort is presented in Table 3.1. The quality of reflection on how molecular-level ideas changed improved significantly over that seen in models written by the same students during the introductory activity ($U(36)=118.5, p=0.026$), illustrating continued improvement as students gained further experience with writing refined models.

3.6 Outcomes: Model Quality for 2009 Cohort

Because of the small sample sizes and low completion rates in both the 2006 and 2008 cohorts, additional data were collected and analyzed during the fall 2009 semester. Aside from minor revisions to sample models between the 2008 and 2009 implementations, the only changes made in 2009 were to shift individual student review of sample models to a pre-lab homework assignment rather than a during-lab assignment and to ask students to provide written explanations for those ratings. This afforded more time during the laboratory period for model writing and for discussion of both sample models and peers' models. In the general chemistry section studied, 78% of students ($n=18$) wrote refined models that were complete enough to be analyzed. The higher during-class completion rate, which was almost certainly a result of making part

of the activity a homework assignment, provides a window into the impact of this activity on a larger cross-section of the student population.

Compared to the 2006 population, the 2009 cohort's models were of significantly higher quality with respect to use of molecular-level language ($U(30)=77.0, p=0.033$) and discussion of model refinements ($U(30)=49.0, p=0.002$), but not in terms of use of experimental evidence ($U(30)=105.0, p=0.115$), to support changes to molecular-level models. Differences between the 2008 and 2009 populations did not approach statistical significance ($p>0.200$ for each category).

3.7 Discussion

Improvement in students' engagement in key thinking processes between analogous model assignments from 2006 to 2008 reflects the gains that are possible as a result of participation in model-critiquing and model-writing activities. Without scaffolding and support for these cognitively complex tasks, it is not surprising that students exhibit relatively poor performance. However, this work allows us to identify characteristics of activities instructors may use to foster greater success at developing scientific explanations that rely on experimental evidence, both in inquiry-based laboratories and in other settings.

One primary difference between the 2006 and 2008 implementations of the MORE introductory activity was the inclusion of sample models. Although the 2008 cohort had, as measured by placement exam scores, a weaker chemistry background, students in that cohort were able to apply knowledge gained from studying sample models of an entirely different system (ethanol and water mixing) to writing models of the candle system, ultimately producing higher-quality models than the 2006 cohort,

and demonstrating greater evidence of thinking processes that are associated with a deeper understanding of the chemistry content and ability to apply it in new contexts.

Because multiple changes were made to the activities between the 2006 and 2008 implementations – the introduction of both sample models and critiquing activities – it is not possible to disentangle the separate effects of each on students' learning. However, because students, drawing on both past experience and the first weeks of a chemistry course, have already been exposed to many examples of molecular-level language that they can emulate, it may be that inclusion of critiquing activities enhanced students' ability to accurately assess the appropriate level of molecular-level language to use in their models. To accomplish any complex task, an individual must first have a reliable and correct understanding of what the task demands as well as the ability to accurately monitor one's progress towards completion. This metacognitive monitoring skill may be enhanced by the process of critically evaluating the completeness of their own and others' models, and then having those evaluations benchmarked to the instructor's and curriculum's standards during class discussion. Indeed, activities involving providing justification for critiques have been studied in the field of cognitive psychology as a means for improving the accuracy of self-assessments of performance.⁵⁰ The effectiveness of even relatively brief exercises in reviewing and critiquing speaks to their potential value as tools to help students learn to construct and support their own scientific explanations.

The importance of students' ability to monitor their own performance, as expressed within critiquing activities, is an area ripe for further study, particularly with respect to the cognitively complex tasks demanded in inquiry-based curricula. Inspired in part by the success of this activity, a second week of activities based on similar

principles, focusing on writing and reviewing scientific explanations, has been incorporated into the MORE general chemistry laboratory curriculum.

Within this activity, which is presented in Appendix E and which follows the first regular laboratory module of the semester, students begin by writing what amounts to an initial model (in the form of an email to a friend) of how to write a MORE refined model. Through scaffolded activities that include consideration of their instructor's written feedback, evaluations of peers' models, delivering specific feedback to peers, comparing rankings of peers' models with each other, conducting additional experiments related to the burning candle system, rewriting the "molecular-level" portion of their candle models, and considering the need to present a convincing case in their models when discussing evidence, students are prompted to develop a more sophisticated understanding of how to write their refined models. To date, the effectiveness of this activity has not been formally assessed. However, the activity culminates in a post-laboratory assignment in which students are asked to rewrite and revise their first full refined model of the semester, and instructors have anecdotally reported that students' models demonstrate substantial improvement in quality following this second activity, and that students express a clearer understanding of what to include in their models and how to go about writing them.

Chapter 4: Student Self-Assessments and Relationships with Model Quality and Instructional Conditions

In Chapter 3, I suggested processing information during introductory activities focusing on model critiquing contributed to students' ability to self-assess their own work against the benchmarks set in example models and during group and class discussion. Based on these experiences, students were able to improve their performance at crafting models that exhibited the features desired by their instructors and the developers of the MORE laboratory curriculum.

While peer and self-critiquing were integral parts of the introductory activities, the previous chapter did not examine the content of those critiques. Were students' assessments of sample models or their own models accurate? Did they improve with time? How did they vary between students? Was the accuracy of these assessments related in any way to students' performance – that is, the quality and completeness of students' written models?

It is this latter connection – the relationship between self-assessment accuracy and performance – that I sought in particular to study in more detail in the work described in this chapter. In capturing student, instructor, and expert evaluations of student model quality in several laboratory sections, we also gained an opportunity to examine the relationships between instructional conditions and students' self-assessment. Were there characteristics of instructors' feedback and grading that influenced students' judgments of their own work or the quality of the models they produced? A semi-controlled study attempted to control the written feedback received

by students. However, other factors – such as instructor, grades assigned, and, of course, the quality of the models that students themselves produced, also may have factored in to students' self-assessments.

In this chapter, I first review some of the background on quantitative student self-assessments, particularly from the field of cognitive psychology, then describe how I collected both student self-assessments and made judgments of model quality, in addition to ratings from students' instructors in the form of grades. I then discuss relationships between these self-assessments (self-ratings), the grades assigned by instructors (TA ratings), and the quality of the models that students produce (measured by expert ratings). In particular, I seek to address the questions:

How accurately do students evaluate the quality of their models? What factors (instructional or otherwise) are related to students' self-assessment of their performance on key model-writing tasks? How does students' accuracy at evaluating the quality of their own and other's work relate to the quality of the models that they produce?

4.1 Background

Self-assessment, as a facet of metacognition, has been explored within a number of related sub-disciplines – in science education, educational psychology, and cognitive psychology – and these various efforts have not always made contact with each other. Before describing the study, some background on past findings concerned with self-assessment is in order.

For decades, researchers in cognitive psychology have studied students' assessments of their own performance in the controlled setting of the research

laboratory, but, over the past decade or so, these studies have begun to extend beyond the cognitive psychology laboratory into the classroom. In one of the first significant studies to do so, Hacker et al⁵¹, asked students in an educational psychology course to provide predictions and postdictions of their performance on course exams – that is, to estimate their grade on an exam both before and after attempting it. These exams consisted entirely of multiple-choice questions, largely from the test bank provided by the publisher of the courses' textbook. The researchers found that the students who received the highest scores on exams were also those who predicted their scores most accurately; top performers had the smallest differences between predicted or postdicted scores and their actual scores

There may be aspects of students' self-assessments that are related to academic or intellectual performance generally. Indeed, it has been noted that accurate self-evaluation is among the general critical thinking skills that higher education should foster^{52,52} and this is borne out by some experimental data. For instance, student's metacognitive monitoring accuracy for a word memorization list task has been found to be correlated to college GPA.⁵³ In the field of chemistry, the degree of overconfidence students exhibit in their performance on a general chemistry course pretest has been seen to be a predictor of whether a student will fail the course – and, in fact, an even better predictor than actual performance on that pretest.⁵⁴

However, metacognitive judgments have proven relatively difficult to influence, and the results of studies that attempt to do so have been somewhat conflicting. Bol et al.⁵⁵ found that repeated practice at making metacognitive judgments of exam performance on successive course exams did not improve the accuracy of those judgments. By contrast, Keleman et al.⁵⁶ found that students' metacognitive monitoring

for foreign language words improved after repeated sessions of being asked to make metacognitive judgments, but only for high performers.

In some settings, feedback has been seen to improve the accuracy of metacognitive judgments. Labuhn et al.⁵⁷ found that fifth-grade students who received feedback on their performance on math problems made more accurate self-assessments of their performance than those who did not. Miller and Geraci⁵⁸ tested the combined effects of feedback and incentives for accuracy and found improvements in metacognitive calibration on exams for low-performing students.

There are also indications that engagement in metacognitive monitoring may lead to improved performance, as we suggest may have been happening during the introductory lab activities described in Chapter 2. Rickey and Stacy^{11c} described how metacognition generally figures into successful problem-solving in chemistry. In the realm of cognitive psychology, Thiede et al.⁵⁹ asked students to generate metacognitive monitoring statements from written texts either immediately or after a delay, and found that students who did so after a delay were more accurate in their metacognitive monitoring. Furthermore, those students in the delayed-monitoring condition engaged in more effective studying when given the opportunity to further review those texts than those in the immediate-monitoring condition, suggesting that improved metacognitive monitoring may lead to improved performance.

Outside the laboratory context, many recent studies on students' metacognitive judgments in the classroom have understandably focused on exams^{54-55, 58, 60} as these data are easy to both collect and process. It is less common to find study of quantitative self-assessment in the context of more complex tasks, particularly in science writing. However, it is precisely because of the challenges students face in these tasks –

particularly assessing their ability at engaging in key scientific thinking processes, such as supporting ideas with reference to evidence – that makes them of interest. Liang & Tsai⁶¹ report, in students' and peers' evaluations of science writing where students were evaluated on broad categories such as knowledge, suitability, correctness, and creativity, that students' evaluations of their peers were generally correlated with experts' ratings, whereas students' evaluations of themselves were generally not. Similar contrasts between self and peer-evaluation for less subjective tasks have also been seen.⁶² Interestingly, in Liang and Tsai's study, the only significant correlations between expert and self-evaluations occurred within the first of three assignments evaluated. Papinczak, Young, Groves, & Haynes⁶³ found only modest correlations between medical students' self-assessments and those of their instructors for performance on problem-based learning modules, which included self-assessment of critical analysis, including basing conclusions on evidence. Although there is significant work studying self-assessment among students in the medical fields, in which evidence-based explanations and reasoning are an important part of professional practice, there does not appear to be examination of self-assessment of science writing in a laboratory context, where developing and writing explanations based on evidence (as described in the previous chapter) has become increasingly emphasized.

4.2 Study Population and Context

The study population for the work described in this chapter and in the following chapter was drawn from five laboratory sections of a chemistry laboratory course taught at Colorado State University in the fall semester of 2008. These sections – the same sections that contribute to the 2008 sample described in Chapter 3 – were from the

first-semester general chemistry course laboratory for science and engineering majors. Within these sections, 97 students consented to participate in the study and have their coursework collected and copied for research purposes.

Demographic summary. Students in this population were overwhelmingly first-year college students; of the 95 consenting students who completed a pre-course survey, 80% reported that they were first-year undergraduates, 12% identified as sophomores, and 4% as juniors. Students overwhelmingly reported college majors related to biological sciences and engineering. The most commonly-reported majors were biological science (15), biochemistry (9), civil engineering (9), zoology (9), mechanical engineering (7), and biomedical science (6). Another 18 students' majors were undeclared, with 6 of these expressing an intention towards a health/life sciences major and 7 expressing an intention towards a math, physical science, or engineering major. Students were almost evenly divided by gender, with 52% self-reporting as female and 48% identifying themselves as male. While the overwhelming majority of students (84%) reported having one year of high school chemistry, 12% reported having 2 years of high school chemistry coursework, while 4% reported having no high school chemistry background. Average self-reported ACT scores were 27 for the Reading/Verbal section, 25 for the English section, and 27 for the Math section. For comparison, the 25th and 75th percentiles of ACT scores for entering first-year students at Colorado State are 21 and 27, respectively, for the English section and 22 and 27 for the Math section.

Instructors. Of the three laboratory instructors (hereafter referred to as TA 1, TA 2, and TA 3), all had multiple semesters of prior experience both teaching laboratory courses using the MORE Thinking Frame and developing laboratory activities. TA 1

taught one laboratory section, while TAs 2 and 3 taught two sections each. Aside from instructions regarding written feedback delivered to students, discussed in the next chapter, the instructors were not directed to alter their teaching style in any way for the study.

Curriculum. The outline for the semester's curriculum is presented in Table 4.1. After an introductory week reviewing the syllabus and safety-related preliminaries, students completed the "Learning to Construct and Refine Models" activity (discussed in detail in Chapter 3) during the second week. The remainder of the semester's activities consisted of five multi-week laboratory modules employing the Model-Observe-Reflect-Explain Thinking Frame. Each module focused on a specific set of general chemistry topics, also identified in the table. For convenient reference, these modules will be referred to as "Spectroscopy," "Smells," "Dissolution," "Nanoparticles," and "Thermochemistry" throughout the remainder of this dissertation.

This study focuses primarily on the first, third, and fifth of these modules – Spectroscopy, Dissolution, and Thermochemistry – for several reasons. First, taken together, these three modules offer windows on students' performance at the beginning, middle, and end of the semester. Second, each of these modules focused on a clearly-defined chemical system or systems – the behavior of atoms within a discharge tube in Spectroscopy, the behavior of salt and sugar dissolved in water in Dissolution, and acid-base and combustion reactions in "Thermochemistry." (The initial and refined model assignments for all 5 laboratory modules are included in Appendix F.) Third, all of the laboratory modules had been implemented with students at Colorado State and other institutions for multiple years and refined after study by project team members.

Table 4.1. Schedule for the Fall 2008 semester of MORE laboratory modules at Colorado State University. Stars mark the weeks in which the InSPRE (In-Class Self and Peer-Reviewing Exercise) activity was implemented.

Week	Activity	General Chemistry Topics
1	Introduction, Syllabus, and Safety Information	
2	LABOR DAY – NO CLASS	
3	Learning to Construct and Refine Models	
4	Spectroscopy, Week 1	absorption and emission of light, Beer's law
5*	Spectroscopy, Week 2	
6	The Chemistry of Smell: How Does the Nose Know? Week 1 ⁶⁴	molecular shape
7*	The Chemistry of Smell: How Does the Nose Know? Week 2	
8	What Happens When Substances are Added to Water? Week 1 ²⁹	aqueous solutions, behavior of electrolytes and non-electrolytes, precipitation reactions
9*	What Happens When Substances are Added to Water? Week 2 ²⁹	
10	What Happens When Substances are Added to Water? Week 3	
11	Exploring Gold Nanoparticles, Week 1 ³²	nanoscale science, Tyndall effect
12*	Exploring Gold Nanoparticles, Week 2 ³²	
13	Laboratory Practical Examination	
14	FALL BREAK – NO CLASS	
15	Chemical Change and Energy: What Fuel Makes the Best Energy Source? Week 1	calorimetry, reaction thermodynamics, bond energies
16*	Chemical Change and Energy: What Fuel Makes the Best Energy Source? Week 2	

To study relationships between three different evaluations of students' models – their self-ratings, their TAs ratings, and a common standard of expert ratings – self-

ratings were captured using an activity called “InSPRE,” (In-Class Self and Peer Reviewing Exercise) described in Section 4.4. TA ratings are the most straightforward, as each TA assigned grades using a grading rubric sheet provided with each model (for the purposes of parallel reference, I will refer to these throughout as TA “ratings,” as grades are a form of rating). An example of this is shown as Figure 4.1 .Capturing these for analysis simply entailed transcribing the TAs numerical assessments for each aspect of the students’ models. The expert rating coding scheme, which was applied to all models analyzed to provide a common standard, is described below.

4.3 “Expert” Model Quality Ratings

Developing a reliable standard by which the quality of models can be judged is not a trivial task. Instructors who are experienced at teaching MORE-based laboratory modules have described using grading methods that range from highly-structured, multi-page rubrics detailing precisely how many points are to be awarded or deducted for specific types of statements to holistic schemes that are not readily translated into well-defined rules. For the purposes of this study, though, it was necessary to have a method of evaluating the quality of student work that (a) could be consistently applied across all sections and all students; (b) could readily be used by other researchers to determine inter-rater reliability, and (c) operated as similarly as possible across model assignments to look at relationships between student performance and feedback on parallel tasks in different models.

Present your refined model (20)			
Describes macroscopic model for salt solution	Not present (0)	Present but incomplete (____)	Present and complete (4)
Describes macroscopic model for sugar solution	Not present (0)	Present but incomplete (____)	Present and complete (4)
Describes molecular-level model for salt solution	Not present (0)	Present but incomplete (____)	Present and complete (4)
Describes molecular-level model for sugar solution	Not present (0)	Present but incomplete (____)	Present and complete (4)
Molecular-level model of salt solution is consistent with experimental evidence	Not consistent (0)		Consistent (2)
Molecular-level model of sugar solution is consistent with experimental evidence	Not consistent (0)		Consistent (2)
Explain why your model has changed (36)			
Identifies key changes to macroscopic model of salt solution	Not present (0)	Present but incomplete (____)	Present and complete (6)
Identifies key changes to macroscopic model of sugar solution	Not present (0)	Present but incomplete (____)	Present and complete (6)
Identifies key changes to molecular-level model of salt solution	Not present (0)	Present but incomplete (____)	Present and complete (6)
Identifies key changes to molecular-level model of sugar solution	Not present (0)	Present but incomplete (____)	Present and complete (6)
Uses specific experimental evidence to explain changes to molecular-level model of salt solution	Not present (0)	Present but incomplete (____)	Present and complete (6)
Uses specific experimental evidence to explain changes to molecular-level model of sugar solution	Not present (0)	Present but incomplete (____)	Present and complete (6)

Figure 4.1. A portion of the grading rubric (used by TAs) for the dissolution laboratory module pertaining to model presentation, identification of model changes, and use of evidence to explain changes.

The coding methodology focused on five key aspects of students' models. As they appear in students' self- and peer-rating rubrics, they are (with abbreviations for them used throughout this chapter given in parentheses):

- “Describes macroscopic model.” (Macroscopic Model)
- “Describes molecular-level model.” (Molecular Model)
- “Identifies key changes to macroscopic model.” (Changes to Macroscopic)
- “Identifies key changes to molecular-level model.” (Changes to Molecular)
- “Uses specific experimental evidence to explain changes to molecular-level model.” (Use of Evidence)

These categories are identified in each of the grading rubrics (an example is shown earlier in Figure 4.1), although some rubrics divide them into multiple subsections, adding specifics relevant to the chemical system in question (e.g. “Describes molecular-level model for salt solution,” “Identifies key changes to macroscopic model of reaction of $\text{CH}_3\text{CH}_2\text{OH}$ with oxygen gas,”). Despite the terminology “identifies key changes,” it should be remembered that this portion of the model asks students to describe the status of each key idea within their refined model as compared to their initial model and to identify whether each key idea has changed or remained the same. Also, although students are asked to describe how both their macroscopic and molecular-level models have changed, they are only asked to provide evidence to support changes to their molecular-level model, ideally making a connection between their macroscopic evidence and their refined model's molecular-level ideas in doing so.

As described in Chapter 2, some aspects of models are contingent upon the presence of other portions of the model. For instance, students are asked to describe

multiple aspects of a system from the molecular-level perspective. For each of these aspects of the system, the student should describe how their ideas have changed (or not) from their initial model. Finally, for each of aspects for which students discuss whether and how their model has changed, the student should support keeping or changing that aspect with reference to specific experimental evidence.

For the presentation of model ideas, completeness could simply be judged by counting how many aspects of the system were present as a fraction of the number of different aspects that were asked about in the assignment. However, evaluating the completeness of a students' description of changes is less straightforward. If a student failed to describe an aspect of the molecular-level model (for example, sugar's molecular-level structure before being added to water), then they could not describe how that portion of their model had changed. Instructors noted that taking off points within discussion of changes for failing to describe how an absent part of the model had changed would constitute "double jeopardy," and one omission could propagate through several aspects of the model. Therefore, to the extent allowed by the rubric, instructors reported that they took this into account in their grading – evaluating the completeness of discussion of molecular-level changes, for instance, based on the molecular-level ideas that were actually presented by students.

It was not always possible for instructors to avoid "double jeopardy." For example, within the Dissolution module, discussion related to salt dissolving in water was a separate category on the rubric from discussion related to sugar dissolving in water (multiple systems discussed within the same model – an acid-base reaction and a combustion reaction – were also split apart on the grading rubric for the Thermochemistry module). If discussion of sugar dissolving was omitted entirely, then

there could not be any discussion of changes to that part of the model or evidence supporting those changes, so students would receive no credit for all three categories on the rubric. However, if the student had only omitted part of the requested discussion of sugar, then they might be deducted points for not completely presenting the requested model, but could still be awarded full credit for adequately describing how the sugar-related ideas they presented changed and what evidence supported those changes.

The InSPRE activity which solicited students' ratings of the quality of models also did not ask students to provide separate ratings for discussions of multiple systems within the same model, even when the grading rubric did; the model quality coding scheme also does not make these distinctions. In developing the InSPRE activity and in developing a coding scheme to characterize model quality, I did not use the subdivisions of rubric categories, as I was primarily interested in how students' self-assessments were related to other aspects of their performance for each type of task, and I wanted to be able to compare students' self-assessments for parallel tasks on successive model assignments.

I also considered it important to capture not merely the presence or absence of certain aspects of students' models, but some judgment of their quality, as this also figured prominently in TA's grading decisions (and presumably in students' judgments as well). Therefore, the quality rating for each aspect of the model was calculated using equation 1, as the number of ideas that are discussed completely or specifically (N_{Comp}) plus one half of the number of ideas that are discussed vaguely or incompletely (N_{Vague}), divided by the total number of ideas that "should be" discussed (N_{Theo}):

$$(N_{\text{Comp}} + 0.5 N_{\text{Vague}}) / (N_{\text{Theo}}) \quad (1)$$

For presenting model ideas, whether macroscopic or molecular, the number of ideas that “should be” discussed is defined with reference to the model assignment. For discussing changes to model ideas, the number of ideas that “should be” discussed is determined by the number of ideas that were presented in that student’s written model (whether presented specifically, meriting full credit, or incompletely or vaguely, meriting partial credit). Finally, for identifying evidence to support molecular-level changes, the number of ideas that “should be” discussed is determined by the number of molecular-level ideas for which changes were discussed in that students’ model. Therefore, except for the identity of the original set of ideas that should be presented by students, calculation of a quality rating for each aspect is parallel for each model assignment. Furthermore, if an idea was discussed only vaguely, and not specifically, in a students’ model, it received only half-credit in the count of the “number of ideas that are discussed,” but still counted fully as an idea for which changes should be discussed. For instance, even if a student is vague about their macroscopic ideas for sugar before being added to water, how those ideas have changed is an idea that “should be” discussed.

The delineation of aspects that “should be” discussed within the “presentation” of each model also attempted to preserve parallels between model assignments. For instance, for every refined macroscopic model, the key ideas to be discussed were visual observations of each of the systems in question at each point in time requested by the model assignment, plus the key measurement students took of each system during the laboratory period. (Although measurements were not likely to be predicted by students in initial models, they were commonly discussed and were expected to be described in refined models, and it is only refined models to which this coding scheme was applied.)

For every molecular-level model, the key ideas to be discussed were the molecular-level behavior of each system at each point in time requested in the model assignment.

Therefore, the key ideas sought within each of the three models related to:

- for Spectroscopy: (1 topic in molecular, 2 topics in macro)
 - hydrogen discharge tube after an electrical current is applied
 - observation of the visible light spectrum (macro only)
- for Dissolution: (4 topics in molecular, 6 topics in macro)
 - salt and sugar, each before and after being added to water
 - conductivity measurements of salt and sugar solutions (macro only)
- for Thermochemistry: (6 topics in molecular, 8 topics in macro)
 - an acid/base and a combustion reaction, each before, during and after the reaction
 - measurements of the temperature changes resulting from each reaction (macro only)

For visual descriptions to be considered specific, and not vague or incomplete, they had to include a description of the color and phase of each substance in question. For measurements to be considered specific, they had to contain specific measured quantities (locations of spectral lines, conductivity measurements, temperature readings). For molecular-level descriptions to be considered specific, they had to identify particular molecular-level species and their structure or dynamics in a way that was consistent with a particulate understanding of matter – the same criteria as the distinctions made regarding molecular-level ideas in Chapter 3.

Discussions of changes were considered to be specific if they contained statements about how and whether their ideas had changed that could be evaluated in

terms of accuracy (although the accuracy of these statements was not evaluated as part of quality ratings), and considered to be vague if changes or similarities from the initial model are alluded to, but not described in sufficient detail that they could be evaluated for accuracy.

Use of evidence to support changes was considered specific if specific data points or observations were identified and the relationship with the model refinements was described. If the evidence was only described imprecisely or indirectly (e.g., “the conductivity readings showed...”, “because of the temperature change”) or if the relationship with model refinements is merely mentioned, but not described specifically, then the use of evidence was considered vague.

The following example illustrates how the quality coding scheme was applied. The refined model assignment for the Dissolution module asks the student to describe salt and sugar on the molecular-level both before and after being added to water. Student 08-045 offered the following molecular-level description:

“Before the NaCl is put into H₂O, the compound looks something like this:

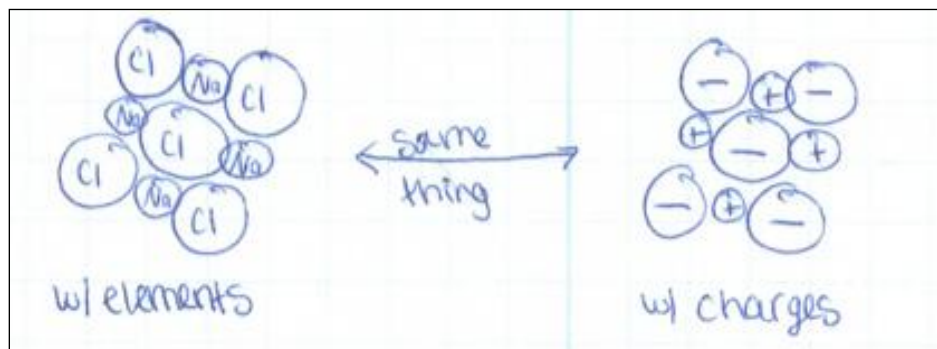


Figure 4.2. Student 08-045’s molecular-level picture of solid NaCl.

The compound is in an ionic lattice. When NaCl is added to H₂O, the electronegativity of H₂O begins to affect the NaCl compound. The ions of NaCl (Na⁺ and Cl⁻) eventually break apart from each other because the attraction between an ion and part of the H₂O molecule is greater. The H₂O molecules kinda ‘section off’ the ions. Notice how the positive H atoms are attracted to the negative Cl⁻ and the negative O is attracted to the positive Na⁺ ion.”

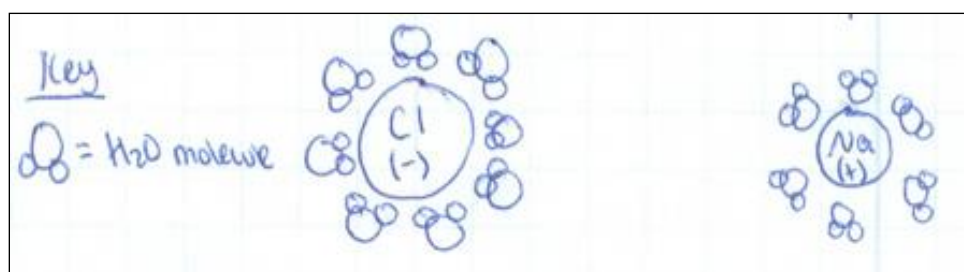


Figure 4.3. Student 08-045’s molecular-level picture of dissolved NaCl.

“NaCl is ionic and so it dissolves like that. Sugar is covalent, so it dissolves differently. A sugar molecule has its own covalent bonds, and it doesn’t break up into ions, even though it does dissolve. [...] Experimental evidence on this was observed in lab, such as the little LED light flashing 11 to 13 times for sucrose and flashing 46-56 times for NaCl. [...] So conductivity, in a way, shows and supports that ionic compounds do indeed get split into ions.”

“Flashing” here is a reference to the conductivity measurements taken by students. The conductivity meters that were used during these experiments indicated conductivity levels using an LED that flashed on and off at a faster rate when the conductivity probe was immersed in higher conductivity solutions. Students were

instructed to take measurements by counting the number of flashes that occurred during a fixed period of time.

Rating the overall quality of this students' molecular-level model (and the related aspects of changes to her molecular-level model and use of evidence to support those changes) begins with identification of the quality of each of the ideas asked about in the assignment:

- Molecular-level ideas of salt before being added to water: The student draws a molecular-level picture of sodium and chloride ions and refers to them existing in an ionic lattice, constituting a specific description.
- Molecular-level ideas of salt after being added to water: Again, a molecular-level picture and detailed textual description of the behavior of “The ions of NaCl (Na⁺ and Cl⁻)” and water molecules constitute a specific description.
- Molecular-level ideas of sugar before being added to water: The student fails to describe sugar before being added to water; ideas about this are absent.
- Molecular-level ideas of sugar after being added to water: Although the student refers to the fact that “sugar molecules” “aren’t broken up,” he or she does not describe what molecular-level species are present in the resulting solution or how they are behaving. Because of the reference to molecular-level particles, this aspect is present, but vague.

Within the overall quality rating, therefore, this student would receive credit for two and a half present ideas out of the four that were expected, resulting in a percentage quality rating of 67%. For description of molecular-level changes, one would then expect

the student to address how the three present ideas (salt before mixing, salt after mixing, and sugar after mixing) changed.

Under the heading “Changes to model,” the student writes:

“In my initial, I said NaCl would dissolve into Na^+ and Cl^- ions. I also said Na^+ would be attracted to O, and Cl^- would be attracted to H. However, my drawing was pretty off. In my refined, I basically explained the same thing, but drew a better and more accurate picture. I also used experimental evidence involving conductivity and the fact that ionic compounds conduct more – 50’s in the glass count) to show/support my guess. For example, NaCl had ~ 46-56 flash times. In my initial for sugar, I only said the molecule would dissolve. I really had no other ideas. In my refined, I explained that conductivity levels are lower because sugar molecules tend to already have plenty of polar bonds, and they just aren’t broken up by H_2O because the H_2O isn’t attractive enough to pull already existing bonds apart. Experimental evidence with this also had to do with conductivity. The sugars had way low conductivity levels (sucrose ~ 11-13 and dextrose ~10-14 flashes).” (underlining in original)

The coder would then evaluate the presence and quality of the three discussions of changes expected:

- Molecular-level ideas of salt before being added to water – The student does not discuss how this portion of their model has developed from their initial model to their refined model; this idea is therefore absent.
- Molecular-level ideas of salt after being added to water – The student identifies that their molecular-level views were the same in their initial model, and specifies changes to their molecular-level picture. Having

clearly identified what the changes and similarities were, this discussion of changes is present and specific.

- Molecular-level ideas of sugar after being added to water – Although the student alludes to ideas in their initial and refined models, they fail to draw any comparison between the two or identify how their ideas have changed – so, although this aspect would be considered present, it is also vague.

With one idea expressed clearly and one vaguely out of the three ideas for which changes should be discussed, the overall quality rating for the “Changes to Molecular-Level” portion of this students’ model would be 50%. Finally, regarding the two ideas for which changes are discussed, the coder would examine the model to see whether or not the student supported each with reference to specific experimental evidence.

- Molecular-level ideas of salt after being added to water – The student says that she “used experimental evidence involving conductivity... to show/support my guess” and even cites specific data points, but she does not describe the relationship between the evidence and the changes to their ideas; this aspect is therefore present but vague.
- Molecular-level ideas of sugar after being added to water – The student cites “low conductivity levels” as supporting what may be a new molecular-level idea in her refined model, but does not appear to offer a reason for the connection between this evidence and the ideas she expressed about the degree to which they are “pulled apart.” This aspect is therefore present, but also vague. (The original coder and the inter-rater coder disagreed as to whether this aspect was vague or specific due to the lack of an explicit

connection between the evidence and the molecular-level ideas referred to.)

Overall, therefore, this student's use of evidence consists of two vague references to evidence in supporting molecular-level ideas, meriting a 50% quality rating.

To establish inter-rater reliability, a second researcher familiar with the MORE Thinking Frame and with coding of student models was provided with 10 student models from the Dissolution module to code using the rating system described above. Rather than calculating a Cohen's kappa, assessing agreement on individual codes, we have calculated correlations between the final quality scores for each rater's review each model aspect in order to establish interrater reliability, because it is only these final quality scores that are included in our later analysis.

For the five aspects, the Spearman rank-order correlation coefficient between the quality ratings from each rater is as follows:

- "Macroscopic Model" – $\rho=0.883$, $p=0.001$
- "Molecular Model" – $\rho=0.540$, $p=0.107$
- "Changes to Macroscopic" – $\rho=0.484$, $p=0.156$
- "Changes to Molecular" – $\rho=0.950$, $p<0.001$
- "Use of Evidence" – $\rho=0.583$, $p=0.077$

Although statistical power for a correlation including only ten pairs of ratings is relatively low, for four of the five model aspects, the correlation coefficient between the two expert raters is above $\rho=0.500$, indicating a strong level of agreement between the two sets of ratings. Similarly high correlations, described later in this chapter, between the grades assigned by two of the three instructors and the quality ratings from the first

researcher, support the validity of the coding scheme as reflecting a strong and consistent relationship between the quality ratings and the way in which instructors graded models. Therefore, in the analyses that follow, these expert quality ratings are taken as a measure of model quality that can be compared between instructors' sections and across all students.

4.4 The InSPRE Activity

In the spring semester of 2007, over a year before the main study data were collected, a colleague attempted to implement Calibrated Peer Review to collect MORE model assignments and have students evaluate each other's work. Calibrated Peer Review is an internet-based system for facilitating incorporation of writing into the classroom, encompassing electronic submission, peer evaluation (including training of peer evaluators), and self-evaluation.⁶⁵ However, several logistical constraints made implementation of CPR challenging for MORE models, including:

- The need for multiple due dates – one for students to submit the written assignment, the second for students to review peers' models – was not always compatible with the fast pace of a laboratory course with refined models due nearly every week.
- The technical complexity of incorporating illustrations within models, both for submission and during model review, within versions of the Calibrated Peer Review software available at the time.
- Concerns about students electronically copying text from their initial models directly into their refined models (or from one refined model in a

laboratory module to the next), which would bypass students' personal reflection on their models and how they had changed.

- The desire to have students complete much of their written work within the hard-copy laboratory notebook.

However, as our interest in capturing students' assessment of their own and their peers' work grew, I worked within the constraints of the laboratory period to develop activity that captured both self-reviews and reviews of peers' work.. This activity was termed InSPRE – the In-Class Self- and Peer-Reviewing Exercise – and was pilot tested during the Fall 2007 semester. In brief, InSPRE asks students to rate the quality of several aspects of both their own model and fellow students' models, using a Likert-type scale.

During its full implementation in the Fall 2008 semester, the InSPRE activity was conducted five times, each at the start of the second laboratory period of a module (weeks 5, 7, 9, 12, and 15) as the first activity of the laboratory period. The first refined model of the module was due at the beginning of the class period and was the model reviewed in each case.

At the start of class, students handed in the assignment(s) due and were given directions for the InSPRE activity and the three-page reviewing packet. A sample reviewing packet (with false names) is provided in Appendix D. Since laboratory notebooks contained carbon copies of written models, students still possessed copies of their written models even though they had already turned in the originals. The InSPRE packet directions instructed them to exchange their laboratory notebooks containing copies of their submitted models with particular other students in the class. After completing one review, students again swapped laboratory notebooks and obtained a

second peers' model to review; after completing that, students then received their laboratory notebooks back and were directed by the activity to re-read and review their own model.

Having students review the first refined model within the InSPRE activity afforded the opportunity for students to both revisit their own model of the chemical system and to consider other students' ideas about those chemical systems while they were still in the midst of the laboratory module that those models addressed. We also anticipated that this activity could help students consider other students' ideas and perspectives on those systems, and that this exchange of ideas might facilitate class discussion of student models.

In order to increase the probability that students would see models of contrasting quality – something we believed to be desirable both for students' development of accurate self-assessments and for our research, students' peer reviewers were not selected at random. Students' grades on the most recent model-writing assignments (those graded since the previous implementation of the InSPRE activity, or, for the first implementation of InSPRE, since the beginning of the term) were calculated, and the students within the class who consented to participate in the research study were divided into "high performing" and "low performing" groups based on these recent grades. Students were then assigned peers to review, ensuring that the two fellow students they reviewed included one member from the "high performing" group and one member from the "low performing" group. Otherwise, assignment was random. A similar procedure was followed for students who did not consent to participate in the research study.

4.5 Analysis of Student Self-Assessments

How do students' self-ratings compare to their instructors' and expert ratings?

Examination of the combined averages for the three types of ratings provided for each aspect of student models – students' self-ratings, TA's ratings in the form of grades, and expert ratings determined by researcher, reported in Table 4.2 – reveal several consistent patterns.

First, students' average self-ratings fall within a relatively small portion of the overall scale compared with the other two types of ratings. Average self-ratings for each aspect within the three models studied all lie between 3.47 and 4.76. Student self-ratings on a Likert scale were converted into percentages for comparison, with the minimum self-rating of "1," indicating that an aspect is "not present," taken to be 0%, and the maximum self-rating of "5," indicating an aspect is "present and complete" taken to be 100%. Intermediate ratings of "2," "3," and "4," would therefore translate to 25%, 50%, and 75%, respectively. When converted to percentages, students' average self-ratings for each aspect translate to values between 64.8% and 94.1%, with all but one aspect averaging above 80%. Most students, therefore, give themselves passing grades on every aspect of every model.

TA's average ratings are generally lower, and the average ratings for the different model aspects cover a much broader range – from 34.7% to 93.8%. Expert ratings tend to be the lowest, with averages ranging from 17.7% to 66.4%. When compared for individual model aspects, expert ratings are significantly lower than student self-ratings in all 15 categories shown in Table 4.2 (all $p < 0.001$, paired samples t-test), and are also lower than TA ratings in all 15 categories (all $p < 0.004$, paired samples t-tests).

Comparisons between student self-ratings and TA grades are of particular interest, though, as the TA ratings (in the form of grades) were a primary source of student feedback during the term. Students did not have had access to the expert ratings. For the majority of ratings (with the exception of the molecular-level model in Spectroscopy and the macroscopic model in Thermochemistry), students' self-ratings are higher than TA ratings, often significantly so; p values for all comparisons between TA and student ratings are given in Table 4.2. The fact that student and TA ratings are as close as they are may be an artifact of both students tailoring their evaluations (intentionally or not) to the grading scale often used in academic settings, in which the vast majority of grades awarded tend to lie in the top 30% of a 0-100% scale, corresponding to "C" grades or higher. As can be seen in Tables 4.2 and 4.3, most (but not all) of the TA grades delivered to students within in this study also fall into this range at the top of the grading scale, but this is not true for every TA and every model aspect.

In every set of models studied, students' ratings for use of evidence are significantly higher than TA ratings ($p < 0.013$ for all cases). Additionally, in the later modules, Dissolution and Thermochemistry, the two out of the five aspects that correspond most directly to aspects identified as important for transfer – changes to molecular model, and use of supporting evidence – are significantly different between the two ratings ($p < 0.028$ for all cases).

How do ratings vary as a function of model aspect?

In addition to being the aspect on which students most consistently and significantly overestimate their performance, use of supporting evidence is also the

aspect of models on which students perform the most poorly. Within both expert and TA ratings, ratings for use of evidence are lower than all other categories within the same model ($p < 0.004$, paired samples t-test, for all comparisons within expert ratings; $p < 0.013$, paired samples t-test for all comparisons within TA ratings). Students' self-ratings also reflect similar patterns within the Spectroscopy module ($p < 0.001$ for all comparisons with evidence category) and the Dissolution module ($p < 0.002$, for all comparisons with the evidence category). Within the thermochemistry module, only description of molecular-level changes is rated lower than use of evidence by students, although this difference is not statistically significant ($p = 0.849$); students rate use of evidence as lower quality than all other aspects in this model ($p < 0.035$).

How do ratings vary throughout the semester?

Students' perceptions of the quality of their own work appear to change in a way that expert or TA ratings do not. Over the course of the semester, the range of ratings that students use becomes substantially truncated. Within in the Spectroscopy module, the difference between the averages for the highest and lowest-rated of the five categories spans 1.12 points (on the scale of 1 to 5), by the Dissolution module, this range narrows to 0.53 points, and by Thermochemistry, the range has narrowed further still to 0.24 points.

Table 4.2. Average expert ratings, TA ratings, and student self-ratings for each aspect of the three models studied, the equivalent percentage ratings for the scales not originally expressed in percentages, and a statistical comparison between the TA and student ratings (paired-samples t-test).

Model and Aspect	Average Expert Ratings	Average TA Ratings, % equivalent (Raw Score)	Average Self-Ratings, % equivalent (Raw Score)	TA vs. Student Ratings, t-test
Spectroscopy:				
Macroscopic	47.7%	87.9% (3.51/4)	89.7% (4.59)	$p=0.568,$ $t=0.574$
Molecular	63.9%	92.6% (3.71/4)	86.8% (4.47)	$p=0.017,$ $t=-0.244$
Changes to Macro	52.9%	78.8% (9.46/12)	82.0% (4.28)	$p=0.420,$ $t=0.811$
Changes to Molec	49.3%	71.7% (8.60/12)	77.6% (4.10)	$p=0.137,$ $t=1.505$
Use of Evidence	17.7%	47.7% (5.72/12)	61.8% (3.47)	$p=0.013,$ $t=2.561$
Dissolution:				
Macroscopic	66.4%	92.3% (7.38/8)	94.1% (4.76)	$p=0.436,$ $t=0.783$
Molecular	45.6%	85.8% (6.87/8)	92.1% (4.68)	$p=0.006,$ $t=2.822$
Changes to Macro	49.4%	80.8% (9.69/12)	88.2% (4.53)	$p=0.045,$ $t=2.048$
Changes to Molec	45.4%	69.5% (8.33/12)	88.2% (4.53)	$p<0.001,$ $t=4.637$
Use of Evidence	28.7%	42.4% (5.09/12)	80.5% (4.22)	$p<0.001,$ $t=7.460$
Thermochemistry:				
Macroscopic	63.0%	89.2% (7.13/8)	89.0% (4.56)	$p=0.932,$ $t=-0.085$
Molecular	38.6%	82.7% (6.62/8)	89.2% (4.57)	$p=0.028,$ $t=2.247$
Changes to Macro	50.9%	85.0% (10.21/12)	88.6% (4.54)	$p=0.182,$ $t=1.347$
Changes to Molec	42.1%	71.8% (8.62/12)	83.1% (4.32)	$p=0.006,$ $t=2.822$
Use of Evidence	26.2%	59.3% (7.12/12)	83.5 % (4.33)	$p<0.001,$ $t=4.614$

Along with this narrowing of the range of average ratings, a gradual increase in student self-ratings can be seen in the significant or near-significant increases in student self-ratings between the Spectroscopy and Dissolution modules within every category of ($p < 0.051$ for all).

This increase in students' assessment of the quality of their work comes despite what is arguably an increasing complexity in the assignments, particularly in the systems that students are asked to discuss – from Spectroscopy, where students are presented with a molecular-level description and are asked to describe what happens afterwards, to the dissolving of compounds in Dissolution, to two chemical reactions in Thermochemistry. One might speculate that the quality of students' work does actually improve as students gain experience in model-writing over the term, but there do not appear to be any such trends for TA or expert ratings between models.

What are the relationships between instructors' grades and students' self-assessments?

As described above, it may seem as though the grades provided by instructors do not affect students' self-assessments. Students' – students' self ratings start out generally higher than the grades assigned by instructors and increase, remaining higher than TA ratings throughout the term. However, students are not likely to be aware of the average grade assigned, but to the particular feedback left by their instructor on their own papers. Different patterns of grading and feedback by different instructors may therefore affect students' self-evaluations substantially. The impact of written comments provided by instructors is discussed at length later in this dissertation; here, we examine the numerical grades assigned by instructors.

With three instructors participating in this study, differences in individual grading patterns were anticipated. Past experience has shown that, with different instructors grading different sections of students, measures are often taken (particularly at the end of an academic term) to ameliorate the differences in grading so that students are not penalized simply for enrolling in a section with a stricter grader. (In the case of these laboratory sections and others taught in parallel with those that make up the study population, each TA's average and standard deviation were used to "curve" classes when final grades were assigned.)

The actual differences between instructors' grading were pronounced, as can be seen in Table 4.3, which presents the average grades given in the Spectroscopy model (the first regular refined model of the semester) by each instructor for each model aspect, as well as an overall average for the model as a whole, with all five aspects equally weighted. Here we can see that TA 2 provided substantially higher grades than those given by either other instructor ($p < 0.001$, independent samples t-test). For later model assignments as well, average grades provided by TA 2 are significantly higher than both those provided by TA 1 (Dissolution, $p = 0.022$, Thermochemistry, $p < 0.001$) and those provided by TA 3 (Dissolution, $p = 0.007$; Thermochemistry, $p = 0.009$).

Table 4.3. Average grades provided by the three TAs to their students for the Spectroscopy module, with standard error of the mean reported in parentheses.

	TA 1 Grade (n=16)	TA 2 Grade (n=28)	TA 3 Grade (n=24)
Spectroscopy:			
Macroscopic	65.28% (5.40%)	93.10% (3.01%)	96.00% (1.87%)
Molecular	88.89% (4.62%)	100.00% (0.00%)	88.00% (4.11%)
Changes to Macroscopic	65.74% (9.06%)	89.08% (5.25%)	75.67% (5.25%)
Changes to Molecular	56.48% (7.67%)	90.80% (4.42%)	59.67% (6.41%)
Use of Evidence	17.13% (4.64%)	74.14% (7.70%)	37.67% (7.68%)
Overall	57.40% (4.51%)	89.05% (2.63%)	72.43% (3.28%)
Dissolution:			
Macroscopic	96.09% (3.17%)	93.30% (3.05%)	88.54% (3.68%)
Molecular	89.84% (3.99%)	97.32% (1.97%)	69.79% (3.68%)
Changes to Macroscopic	78.13% (7.05%)	87.20% (5.21%)	75.00% (5.94%)
Changes to Molecular	56.25% (7.08%)	78.87% (4.86%)	67.36% (6.06%)
Use of Evidence	26.56% (7.23%)	48.51% (8.66%)	45.83% (7.56%)
Overall	69.38% (3.74%)	81.04% (3.04%)	69.31% (2.83%)
Thermochemistry:			
Macroscopic	77.24% (3.46%)	99.55% (0.45%)	84.90% (3.68%)
Molecular	68.75% (5.10%)	94.64% (2.68%)	78.13% (4.47%)
Changes to Macroscopic	61.98% (5.94%)	99.70% (0.30%)	83.33% (4.40%)
Changes to Molecular	44.79% (7.83%)	83.93% (5.78%)	75.69% (6.50%)
Use of Evidence	23.44% (6.19%)	72.62% (8.01%)	75.69% (6.50%)
Overall	55.26% (4.26%)	90.09% (2.48%)	77.95% (3.82%)

TA 3 also provided grades that were, on average, higher than those provided by TA 1. Differences in overall averages are significant in the Spectroscopy module ($p=0.009$) and Thermochemistry module ($p<0.001$), but not in the Dissolution module ($p=0.998$). In general, TA 2 provided the highest grades, followed by TA 3; TA 1 provided the lowest grades. These patterns are also observed in the majority of comparisons between individual model aspects.

Did these differences in model grades arise because the actual quality of models was different in each class, or were the TAs actually grading differently? The populations

in each laboratory section were not randomly assigned, but arose due to student selection of laboratory times during their course registration, and are a source of potential disparities in student performance between sections. Similarly, differences in instruction early in the semester (such as during introductory activities) could have resulted in different classes producing different quality models. To make this comparison, the expert ratings for the Spectroscopy model – the first model of the semester – conducted with a consistent methodology across different TA's sections – were compared, and no statistically significant differences were found in combined ratings. Additionally, no significant differences between each set of students were found with respect to their performance on either the math or chemistry subsets of the American Chemical Society's Examinations Institute 1998 Toledo Chemistry Placement Examination administered at the beginning of the semester.

When considered as a whole, average expert ratings also do not significantly differ between TAs at the end of the semester. Total average expert ratings are reported in Figure 4.4 below. Although the differences are not significant, TA 2's students produce the lowest-quality models by the end of the semester; at the same time, they receive significantly higher grades throughout the same term than students in the other laboratory sections studied.

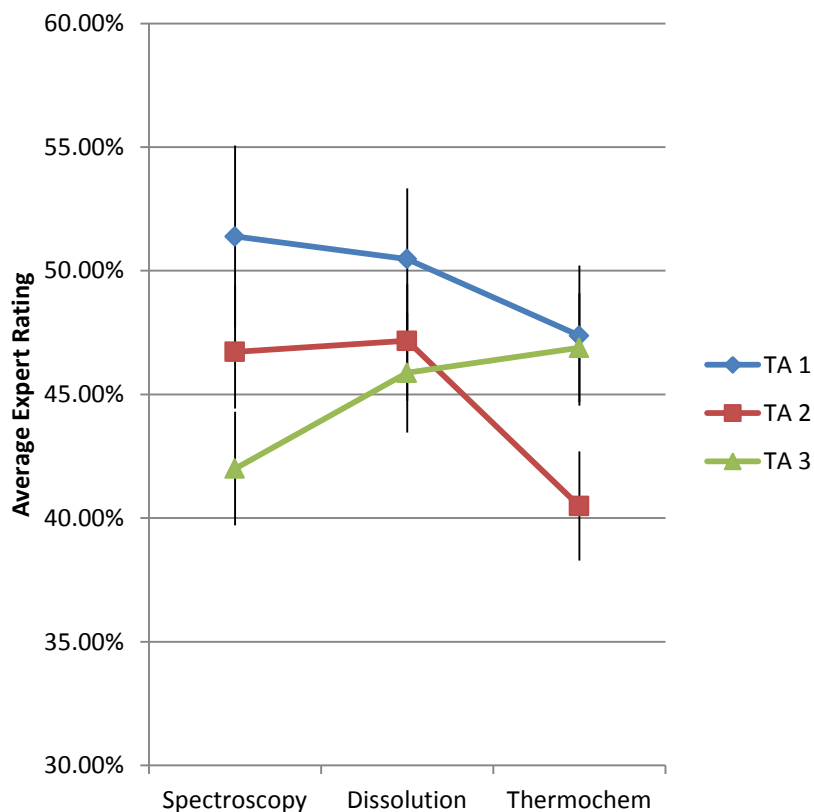


Figure 4.4. Average expert ratings, broken apart by TA, for the major lab modules studied. Error bars are standard error of the mean.

However, one aspect of model quality in which differences do emerge in expert ratings between instructors is in presentation of molecular-level models. Comparisons between TA 1 and TA 2's students reveal no significant difference in expert ratings within the Spectroscopy module ($p=0.453$), but TA 1's students perform better on this aspect, on average, within the Dissolution module ($p=0.026$) and Thermochemistry module ($p=0.003$). A significant difference between expert ratings of TA 1 and TA 3's students' model quality is also apparent in the same category (molecular-level models)

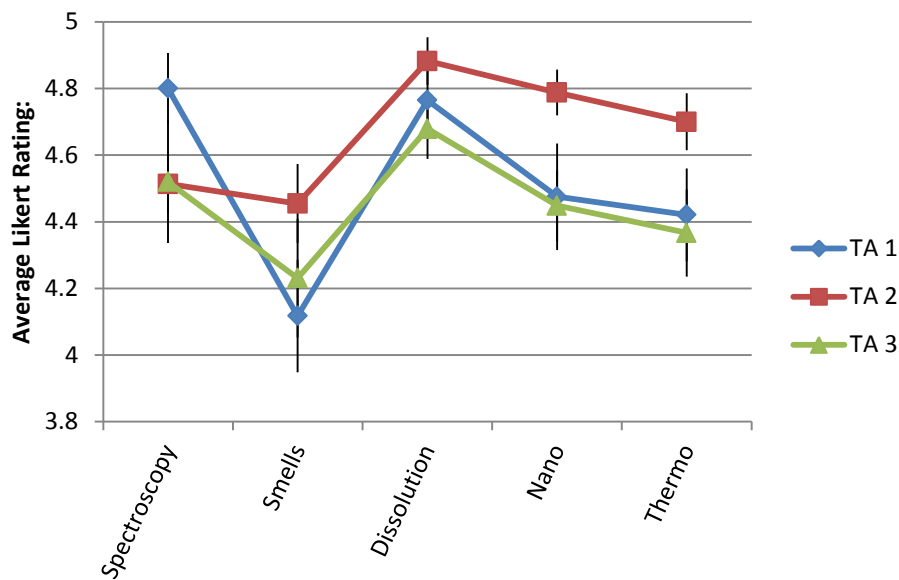
at the beginning of the semester ($p=0.048$) and persists at significant or near-significant levels ($p=0.012$ in Dissolution; $p=0.061$ in Thermochemistry) as the semester continues. However, note that TA 1 awarded students the lowest average grades, so, although this instructor's students may have performed slightly better (based on expert ratings) on one model aspect than the students of other instructor, this instructor can still be regarded as the "strictest" of the three graders.

Did the feedback of higher grades in TA 2's classrooms have an impact on students' self-assessments of their performance? Considered collectively, as early as the first laboratory module, TA 2's students rate themselves higher on average than those of TA 1 ($p=0.020$). This difference is significant not only for the combined ratings, but for the three key aspects of presenting their molecular-level model ($p=0.036$), describing changes to their molecular-level models ($p=0.015$), and using supporting evidence ($p=0.021$). Both the significance and the magnitude of the difference for the combined ratings increase in the subsequent laboratory modules – $p=0.003$, $t=-3.164$ for Dissolution; $p=0.006$ and $t=-2.926$ for Thermochemistry. Similarly, while students of TA 2 and TA 3 begin without showing significant differences in self-evaluations ($p=0.238$, $t=1.196$), the significance and magnitude of the difference, with TA 2's students rating themselves more highly, increase by the Dissolution ($p=0.001$, $t=3.665$) and Thermochemistry modules ($p=0.005$, $t=2.907$).

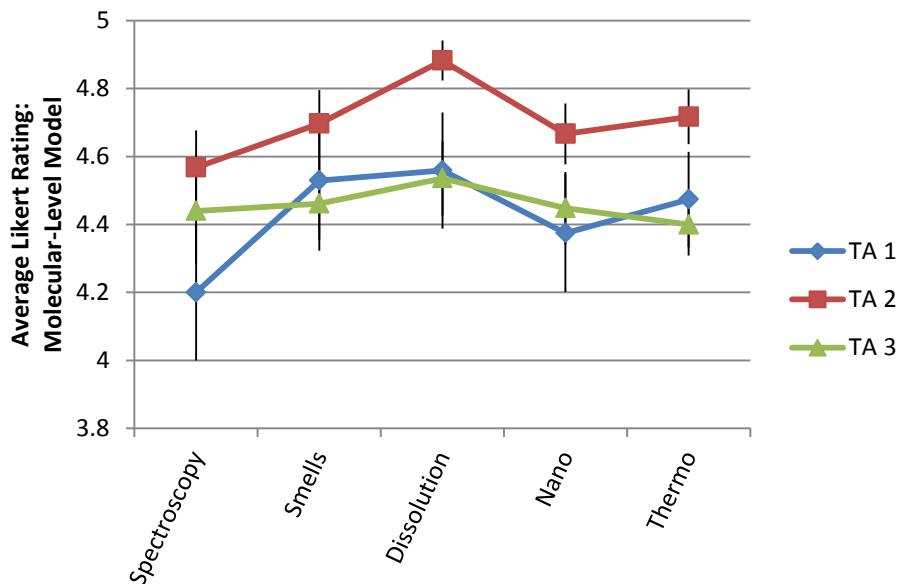
Examination of students' self-ratings within individual categories reveals similar trends. Although TA 2's students do not consistently start out with the highest ratings among the three TAs, their ratings are the highest by the end of the semester, and show general trends of increasing throughout the semester, while TA 1 and TA 3's students' average self-ratings are more mixed. While grades are not the only form of feedback that

could influence students' self-assessments – written feedback on models and verbal feedback within and outside of class are also likely to contribute – grades may be, for students, the most salient and the most easily accessible.

(a) Macroscopic Model:

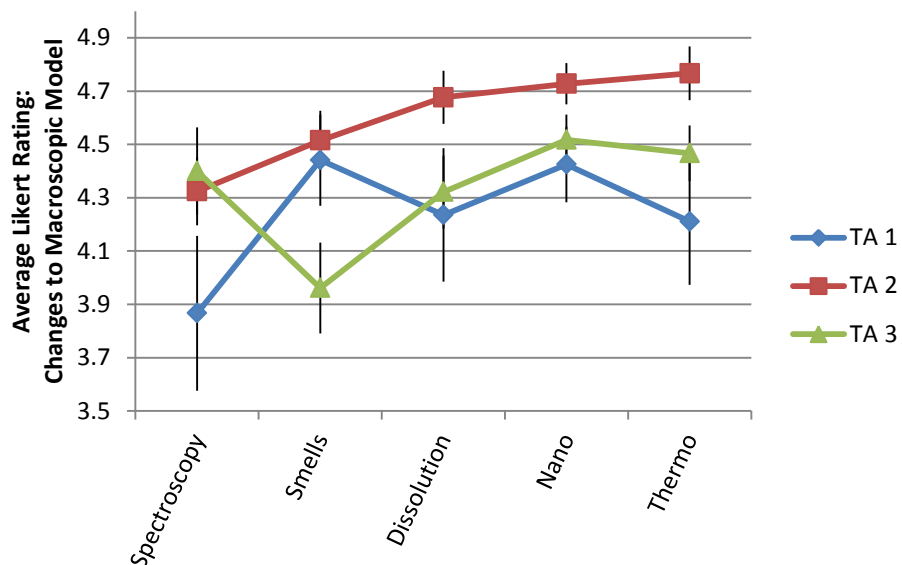


(b) Molecular-level Model:

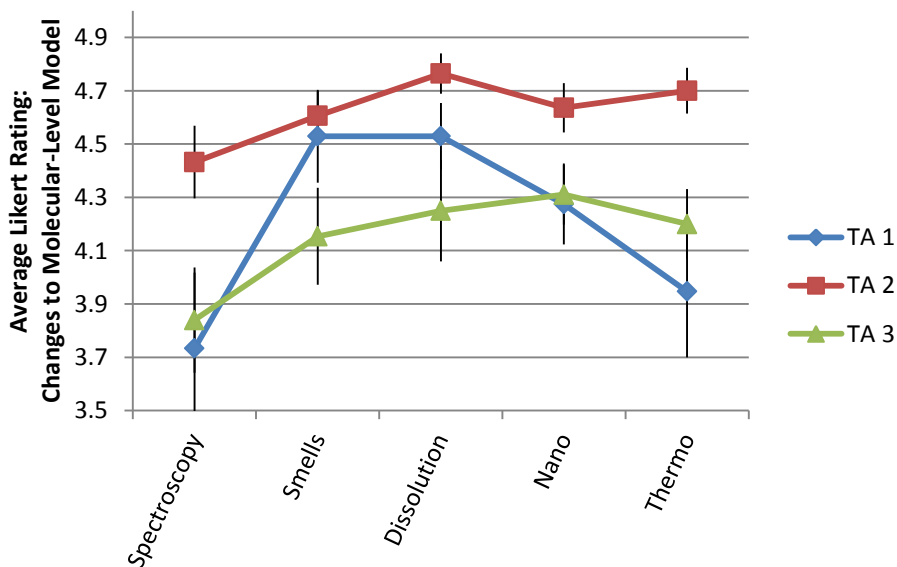


Figures 4.5 (a)-(e). [Above and on following pages.] Average student self-ratings by model aspect, broken apart by TA, for all 5 models reviewed during the Fall 2008 semester.

(c) Changes to Macroscopic:

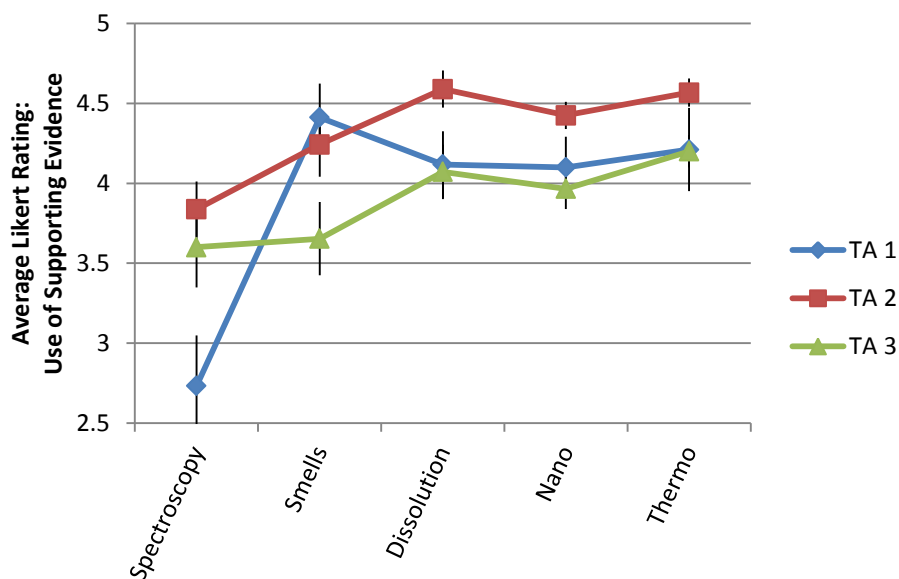


(d) Changes to Molecular:



Figures 4.5 (a)-(e). [Above and on preceding and following pages.] Average student self-ratings by model aspect, broken apart by TA, for all 5 models reviewed during the Fall 2008 semester.

(e) Use of Evidence:



Figures 4.5 (a)-(e). [Above and on preceding pages.] Average student self-ratings by model aspect, broken apart by TA, for all 5 models reviewed during the Fall 2008 semester.

4.6 Calibration of Student Self Assessments

While students' overall self-evaluations may be influenced by factors beyond the quality of the model – such as the grades students receive on various assignments – one would also expect that students' self-evaluations would bear some relationship with the quality of the models or portions of models being evaluated.

The data presented already provide some conflicting indications as to the nature of such relationships. On the one hand, when ratings for different model aspects are compared, the student, expert and TA ratings of performance on individual models aspects all generally parallel each other – students perform better (as judged by both

expert and TA ratings) and judge themselves to be better at describing their models than at describing changes to those models or, particularly, at using evidence to support changes to those models. This suggests that, at least in one respect, students – or at least some subset of the student population – reliably recognize differences in quality. A longitudinal view provides a different impression, though. Students' self-ratings increase over the course of the semester, a trend not supported by TA or expert ratings. The sometimes large divergence between students' self-ratings and TAs ratings – especially when it is observed that student self-ratings seem to mirror TA ratings to some extent – raises questions about whether students' ratings meaningfully discriminate between higher-quality and lower-quality work, and to what degree self-assessment plays a role in students' model-writing.

Do students' provide higher ratings to models that are actually higher quality?

One way of addressing this question is to look at correlation coefficients between students' self-ratings and external (TA and expert) judgments of quality. In this study, I have used non-parametric Spearman correlation coefficients throughout, consistent with the fact that the three rating measures cannot be considered continuous variables, as students and TAs particularly had a constraint range of grade or rating choices (1, 2, 3, etc.) for each model aspect. The top section of Table 4.4 presents Spearman correlations for combined ratings (the average of all 5 aspects for an individual model). Interestingly, while overall student self-ratings for each model are significantly correlated with TA ratings, with correlation coefficients between +0.265 and +0.371 ($p < 0.029$), they are not significantly correlated with expert ratings ($p > 0.124$ for all). Note that the coding scheme used for expert ratings was never made available to

students, so they did not use the expert rating criteria directly to inform their self-assessments.

Table 4.4. Correlation coefficients between combined ratings (the average of all 5 aspects considered) for the three sets of ratings and three models evaluated (top section), and correlations between ratings broken down by model aspect (bottom sections). Statistically significant correlations ($p < 0.050$) are highlighted in bold.

	Spectroscopy	Dissolution	Thermochemistry
Correlations (combined):			
Self-Rating vs. TA Rating	+0.371, $p=0.002$	+0.275, $p=0.023$	+0.265, $p=0.029$
Self-Rating vs. Expert Rating	+0.163, $p=0.184$	+0.142, $p=0.246$	+0.189, $p=0.124$
TA Rating vs. Expert Rating	+0.338, $p=0.005$	+0.488, $p < 0.001$	+0.377, $p=0.002$
Self vs. TA (by aspect):			
Macroscopic Model	+0.026, $p=0.833$	+0.122, $p=0.321$	+0.262, $p=0.031$
Molecular-Level Model	+0.303, $p=0.012$	+0.375, $p=0.002$	+0.133, $p=0.279$
Changes to Macroscopic	+0.248, $p=0.042$	+0.165, $p=0.178$	+0.304, $p=0.012$
Changes to Molecular-Level	+0.363, $p=0.002$	+0.094, $p=0.445$	+0.248, $p=0.041$
Use of Evidence	+0.243, $p=0.046$	+0.116, $p=0.348$	+0.031, $p=0.800$
Self vs. Expert (by aspect):			
Macroscopic Model	+0.051, $p=0.678$	-0.064, $p=0.604$	+0.395, $p=0.001$
Molecular-Level Model	+0.118, $p=0.338$	+0.193, $p=0.115$	-0.058, $p=0.638$
Changes to Macroscopic	+0.173, $p=0.159$	+0.148, $p=0.230$	+0.066, $p=0.593$
Changes to Molecular-Level	+0.242, $p=0.047$	-0.008, $p=0.945$	+0.205, $p=0.094$
Use of Evidence	+0.396, $p=0.001$	+0.141, $p=0.252$	+0.121, $p=0.326$
TA vs. Expert (by aspect):			
Macroscopic Model	+0.179, $p=0.143$	+0.426, $p < 0.001$	+0.292, $p=0.016$
Molecular-Level Model	+0.340, $p=0.005$	+0.343, $p=0.383$	+0.221, $p=0.070$
Changes to Macroscopic	+0.416, $p < 0.001$	+0.383, $p=0.001$	+0.244, $p=0.045$
Changes to Molecular-Level	+0.322, $p=0.007$	+0.377, $p=0.002$	+0.519, $p < 0.001$
Use of Evidence	+0.249, $p=0.041$	+0.432, $p < 0.001$	+0.437, $p < 0.001$

On the other hand, expert ratings reflect many of the standards that the designers of MORE laboratory instruction seek to have students reach. If instructors' ratings did not all align to some extent with the common standard of the expert rating methodology, then it makes less sense to consider students' self-ratings for all three instructors together, particularly when effects of various instructional factors are considered.

Furthermore, although the expert rating scheme is not identical to the coding methodologies used to characterize key thinking processes associated with transfer, this research did inform its construction, and instructors did intend (to varying degrees) to foster these thinking processes through their grading and feedback. It is therefore useful to consider how much instructors diverged from these common standards in their own grading.

Do TA's ratings reflect aspects of performance captured in expert ratings?

Table 4.5 reports Spearman correlation coefficients between TA grades and expert ratings. Both the combined ratings (the average of all 5 aspects), and ratings for individual model aspects are considered within all three models.

When broken apart by TA, correlations between TA grades and expert ratings are higher (the preponderance of individual correlations being above +0.500) than when all TAs are considered together (these correlations, reported above, falling within the range of +0.338 to +0.488). This reflects the fact that, while individual instructors may be reasonably-well calibrated, this relationship is somewhat masked by the fact that instructors may center each of their student populations at different points along the 0-100% grading scale.

One noteworthy finding within these data is that there are indeed differences in the degree to which TA's grades correlate with expert ratings. Both within the aggregate ratings and within individual model aspects, TA 2's grades show the poorest correlation with expert ratings. Among the model aspects examined, TA 2's grades show the lowest correlation with expert ratings for 14 out of 15 cases. The only exception is the macroscopic portion of the Dissolution model, for which TA 1 is the most poorly correlated with expert ratings. Although the correlation coefficient (ρ) between TA 2's ratings and expert ratings is as high as +0.582 in one case (presentation of the macroscopic model in the Dissolution module), it is actually negative for some model aspects. TA 1 and TA 3's grades, by contrast, show many strong correlations with expert ratings, with correlation coefficients as high as +0.854 (TA 1, Dissolution, use of evidence). For TA 1, 12 of 15 individual comparisons on model aspects show significant correlations between TA grades and expert ratings, all of which would be considered strong under Cohen's system for classifying correlation strengths.⁶⁶ (TA 1 graded changes to molecular-level models in a unique manner that may account for poorer correlations in those categories; this is discussed below.) Similarly, for TA 3, 14 of 15 correlations are statistically significant, with 10 of these considered strong. However, for TA 2, only one of the six significant correlations would be considered strong.

Table 4.5. Correlation coefficients between TA’s grades and expert ratings of model quality – both aggregate ratings (top section) and individual aspects for each model studied (bottom 3 sections). *No correlation coefficient could be calculated for these two variables, as TA 2 assigned all students within the population analyzed full credit for this part of their models.

	Spectroscopy	Dissolution	Thermochem.
All Aspects Combined:			
Spectroscopy	+0.500, <i>p</i>=0.049	+0.363, <i>p</i> =0.058	+0.591, <i>p</i>=0.002
Dissolution	+0.830, <i>p</i><0.001	+0.300, <i>p</i> =0.121	+0.581, <i>p</i>=0.003
Thermochemistry	+0.832, <i>p</i><0.001	+0.507, <i>p</i>=0.006	+0.807, <i>p</i><0.001
Spectroscopy:			
Macroscopic Model	+0.708, <i>p</i>=0.002	-0.019, <i>p</i> =0.924	+0.274, <i>p</i> =0.195
Molecular-Level Model	+0.558, <i>p</i>=0.025	N/A*	+0.407, <i>p</i>=0.048
Changes to Macroscopic	+0.519, <i>p</i>=0.040	+0.426, <i>p</i>=0.024	+0.691, <i>p</i><0.001
Changes to Molecular-Level	+0.401, <i>p</i> =0.124	+0.352, <i>p</i> =0.066	+0.459, <i>p</i>=0.024
Use of Evidence	+0.818, <i>p</i><0.001	+0.097, <i>p</i> =0.625	+0.419, <i>p</i>=0.042
Dissolution:			
Macroscopic Model	+0.080, <i>p</i> =0.767	+0.582, <i>p</i>=0.001	+0.408, <i>p</i>=0.048
Molecular-Level Model	+0.662, <i>p</i>=0.005	+0.217, <i>p</i> =0.267	+0.535, <i>p</i>=0.007
Changes to Macroscopic	+0.761, <i>p</i>=0.001	+0.063, <i>p</i> =0.749	+0.665, <i>p</i><0.001
Changes to Molecular-Level	+0.491, <i>p</i> =0.054	+0.247, <i>p</i> =0.205	+0.505, <i>p</i>=0.012
Use of Evidence	+0.854, <i>p</i><0.001	+0.185, <i>p</i> =0.347	+0.580, <i>p</i>=0.003
Thermochemistry:			
Macroscopic Model	+0.717, <i>p</i>=0.002	-0.135, <i>p</i> =0.495	+0.712, <i>p</i><0.001
Molecular-Level Model	+0.797, <i>p</i><0.001	+0.398, <i>p</i>=0.036	+0.695, <i>p</i><0.001
Changes to Macroscopic	+0.644, <i>p</i>=0.007	-0.253, <i>p</i> =0.194	+0.666, <i>p</i><0.001
Changes to Molecular-Level	+0.768, <i>p</i>=0.001	+0.435, <i>p</i>=0.021	+0.634, <i>p</i>=0.001
Use of Evidence	+0.533, <i>p</i>=0.033	+0.442, <i>p</i>=0.019	+0.606, <i>p</i><0.001

Several factors may account for these differences in the degree of correlation between grades and expert ratings. As described earlier, TA 2 delivered substantially higher grades than TA 1 and TA 3, using a smaller portion of the possible grading scale. In fact, TA 2 assigned perfect scores to more than half of students for all but one model

aspect (Dissolution, use of evidence). Additionally, TA 1 and TA 3, in their roles as researchers, were more involved in the development and use of coding schemes to capture the quality of students' model quality than TA1 TA 3 was the primary developer and implementer of the expert coding scheme. Their grading is understandably better correlated with the expert coding results. The strong correlations between their grades and the expert ratings, therefore, offer another piece of evidence in support of the validity of the expert coding scheme as a measure of model quality.

Are there areas where TAs' ratings are better aligned with "expert" ratings?

One might also suspect that, because different instructors seem to have taken differing approaches to grading, that differences might emerge between model aspects or between models in the degree of correlation between TAs' grades and expert ratings – for instance, if one TA's grading processes aligned very closely with the expert rating method for certain aspects, but not for others.

It should be noted that TA 1 deliberately graded discussions of model changes differently from the other instructors. Consistent with research on the MORE Thinking Frame, which shows that student success at transfer is linked with *accurate* reflection upon changes, this instructor reported examining students' initial models when grading this aspect of refined models, and deducting points when students made inaccurate statements about how their models had changed. Interestingly, this instructor also reported, following one implementation of the InSPRE activity, that a student had inquired how they were to evaluate the quality of discussion of model changes without having the initial models available for reference. No students of the other two instructors made such inquiries. It is therefore interesting to observe that, among the

three aspects where this instructor's grades were not strongly correlated with expert ratings, two of these were for discussion of changes to molecular-level models (in Spectroscopy and Dissolution).

Examination of other aspects for all three TAs, however, fails to yield any other discernible patterns in the strengths of correlations. For example, while, for TA 1, grades for use of evidence are most highly correlated with expert ratings in the Spectroscopy and Dissolution modules, they are the most poorly correlated aspect in the Thermochemistry module. For TA 2, none of the model aspects are significantly correlated with expert ratings in more than one module. And for TA 3, although the macroscopic model is the most poorly correlated aspect in the Spectroscopy and Dissolution modules, it is the most highly correlated aspect in the Thermochemistry module.

How well-calibrated are students' judgments?

In this context, calibration is defined as the difference between students' actual performance and their judgments of their own performance. This index of calibration is sometimes referred to as "bias:"⁶⁷

$$\text{Calibration} = \text{predicted score} - \text{actual score} \quad (1)$$

This measure of calibration can therefore have both positive values (if a student predicts a greater score than they actually achieve, reflecting overconfidence) and negative values (reflecting underconfidence). Calibration is sometimes reported as simply a measure of precision, reflecting the absolute value of the difference between predicted and actual score; however, since this study will discuss how various instructional may affect

confidence judgments, the index of calibration above will be used throughout this dissertation, consistent with the recommendations of Schraw (2009).⁶⁷

Because two different forms of more “actual” scores are available within this study – the TA grades and the expert ratings – calibration can be calculated with respect to both – and both are reported in Table 4.6 below. Many of the patterns that can be gleaned from these average calibration scores were discussed in conjunction with average ratings in Table 4.2. For all but two categories, the average difference is positive, reflecting the fact that students’ self-ratings are markedly higher than the TA or expert ratings. It can be seen quite clearly here that calibration is poorest in each model for use of evidence, consistent with the average scores discussed above.

Note that in this data, calibration scores are reported averaged across all students, regardless of instructor. It could be argued that, when reporting average calibration scores and measures that depend on them (as will be done in the following sections), each instructors’ grades are distinct measures, and that such correlations or measure involving calibration should be analyzed and reported separately for each instructors’ population – to some extent, this is done below in later sections with all model aspects averaged together (such as in Tables 4.7, 4.9, and 4.10) to demonstrate that similar behaviors are seen within different instructors’ populations. In other cases (such as Table 4.6 above, and Figures 4.6-4.8 below) the relatively small size of the standard error of the mean offers a good indication that the behaviors described are relatively consistent across the entire student population, and that overall patterns are not merely an artifact of differences between instructors’ grading practices. Most noticeably, the average differences between students’ ratings and TA ratings are smaller for every category and model aspect than the differences between student ratings and

expert ratings. This, combined with the higher correlation between student self-ratings and TA grades, as well as the way in which students' ratings evolve over the semester to reflect differences in TA grades, helps support the conclusion that students base their judgments in part on the feedback provided by their instructors.

Table 4.6. Average calibration (the differences between predicted and actual scores) for each model aspect; calibration calculated with respect to both TA ratings and expert ratings. Standard error of the mean is reported in parentheses.

Model Aspect	Calibration with Respect to TA Ratings	Calibration with Respect to Expert Ratings
Spectroscopy:		
Macroscopic Model	+1.84% (3.20%)	+41.91% (3.37%)
Molecular-Level Model	-5.88% (2.41%)	+22.79% (3.42%)
Changes to Macroscopic	+3.19% (3.93%)	+29.04% (4.27%)
Changes to Molecular-Level	+5.88% (3.91%)	+28.31% (4.12%)
Use of Evidence	+14.09% (5.50%)	+44.12% (4.03%)
Dissolution:		
Macroscopic Model	+1.84% (2.35%)	+27.71% (2.76%)
Molecular-Level Model	+8.27% (2.44%)	+46.84% (2.73%)
Changes to Macroscopic	+7.48% (3.65%)	+38.88% (3.71%)
Changes to Molecular-Level	+18.75% (4.04%)	+42.96% (3.84%)
Use of Evidence	+38.11% (5.11%)	+51.84% (4.04%)
Thermochemistry:		
Macroscopic Model	-0.18% (2.15%)	+26.19% (2.30%)
Molecular-Level Model	+6.43% (2.86%)	+50.54% (3.07%)
Changes to Macroscopic	+3.55% (2.64%)	+37.78% (3.40%)
Changes to Molecular-Level	+11.27% (4.00%)	+46.62% (3.45%)
Use of Evidence	+24.14% (5.23%)	+56.88% (3.74%)

In what circumstances do students' ratings correlate with their TA's ratings?

As students seem to calibrate their judgments of their own performance towards the ratings (grades) provided by their instructors, it makes sense to examine the

correlation between students' self ratings and TA's ratings in more detail, as presented in Table 4.7. There do not appear to be consistent patterns in the correlations between students' self ratings and instructor's grades. However, it is interesting to note that correlations increase in magnitude across the semester among TA 1's student population, becoming statistically significant in the final Thermochemistry module, whereas, among TA 2 and TA 3's students, correlations are statistically significant in the Spectroscopy module at the start of the semester, but not in later modules. It is conceivable that other instructional factors unique to TA 1, such as stricter grading and (as discussed in Chapter 5) a greater quantity of written feedback promoted a greater correlation between TA and student ratings than for other instructors.

Additionally, when broken apart by individual aspect and laboratory, the fewest number of individual aspects (only one out of five) show significant correlations within the Dissolution laboratory, while within both the Spectroscopy and Thermochemistry laboratories, four out of the five aspects considered are significantly correlated. The reasons for this difference between modules are not immediately apparent. Correlation coefficients, however, can be driven by outliers – data points at either the high or low end of the scales which can distort correlation coefficients. Calibration scores come in useful in helping avoid some of these difficulties.

Table 4.7. Correlation coefficients between students' self ratings and TA ratings, broken apart by model aspect (top section) and, for overall correlations, by TA (bottom section).

	Spectroscopy	Dissolution	Thermochemistry
Combined Ratings	0.371, $p=0.002$	0.275, $p=0.023$	0.265, $p=0.029$
Ratings by Aspect:			
Macroscopic Model	+0.011, $p=0.929$	+0.052, $p=0.668$	+0.279, $p=0.021$
Molecular-Level Model	+0.352, $p=0.003$	+0.419, $p<0.001$	+0.121, $p=0.037$
Changes to Macroscopic	+0.308, $p=0.001$	+0.231, $p=0.058$	+0.461, $p<0.001$
Changes to Molecular-Level	+0.395, $p=0.001$	+0.057, $p=0.643$	+0.368, $p=0.002$
Use of Evidence	+0.287, $p=0.018$	+0.159, $p=0.197$	+0.131, $p=0.286$
Ratings by Instructor:			
TA 1	-0.038, $p=0.889$	+0.182, $p=0.499$	+0.579, $p=0.019$
TA 2	+0.469, $p=0.012$	+0.311, $p=0.107$	+0.005, $p=0.979$
TA 3	+0.489, $p=0.015$	+0.197, $p=0.356$	+0.067, $p=0.757$

Furthermore we are also interested in individual differences. How do students who rate themselves highly differ in performance from others, for instance? To examine the distribution of students' ratings, we construct calibration plots, a common way of looking at such data within the field of cognitive psychology.⁶⁸

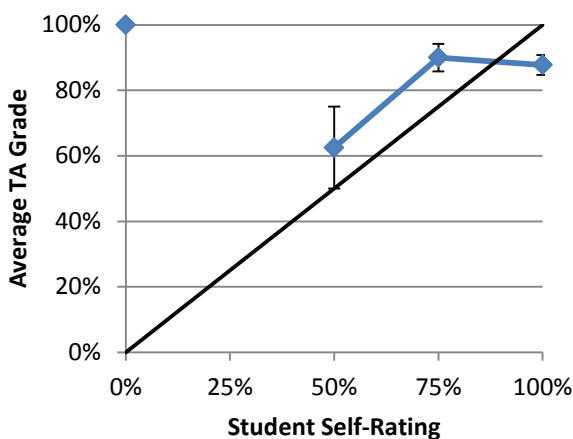
Calibration plots illustrate the average performance on items as a function of subjects' own self-evaluations. For example, for students who rated themselves as a "4" (75%) on an aspect of a model, what was the average rating provided by TAs? If students and TAs were in perfect agreement, such a plot would produce a straight line along $y=x$. Figures 4.6 – 4.8 below present these data for each of the five model aspects. Error bars in these plots reflect the standard error of the mean for that particular set. Data points without error bars reflect cases in which there was only one student rating. The final plot for each set reflects the combined set of ratings for one model.

Points that lie above and to the left of the diagonal reflect students underestimating the grades they receive from their TA; points that lie below and to the right of the diagonal reflect students' overestimating their grades. Obviously, unless students are in perfect agreement with the TA, students' will underestimate their grades at the lowest end of the scale (an estimate of 0% can only be correct or an underestimate) and overestimate their grades at the highest end (100% can only be correct or an overestimate). For many model aspects, no students rated themselves at the lowest end of the scale ("1" or "2"), resulting in a limited range of points.

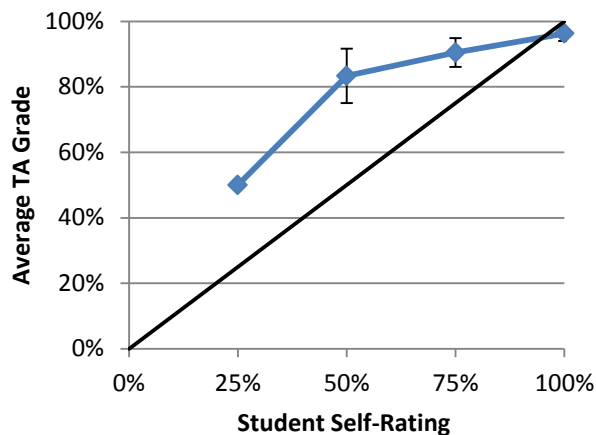
The trends exhibited in the calibration plots are similar to those reported in the averages above – use of evidence shows the greatest degree of overestimation, with discussion of changes showing less overestimation and presentation of both models the least. However, students' ratings of the quality of their macroscopic models are often underestimated as compared to their TAs, a conclusion that is not apparent from the calibration averages reported above. Several model aspects for which the average values reflect overall overestimation (Spectroscopy, changes to molecular; Thermochemistry, changes to molecular) actually exhibit reasonable alignment with the calibration curve, particularly over the "intermediate" range of student self-ratings (25%-75%).

In fact, when histograms for the degree of calibration for individual model aspects are constructed (Figures 4.9 a-c), it becomes clear that students are actually reasonably well calibrated to their instructors' grades. These graphs illustrate the distribution of differences between students' self-evaluations and instructors' grades. Because students' self-evaluations are constrained by the InSPRE activity to fall at one of 5 discrete values for each individual aspect (each placed 25% apart on a 0-100% scale), bins of a width of 25% of the calibration value were used in these histograms.

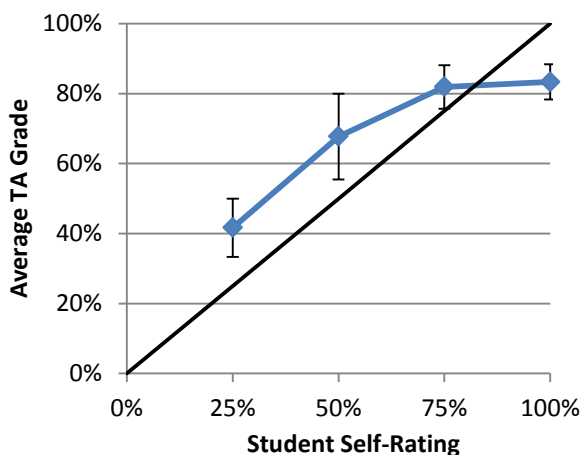
4.6(a) Spectroscopy, Macroscopic Model:



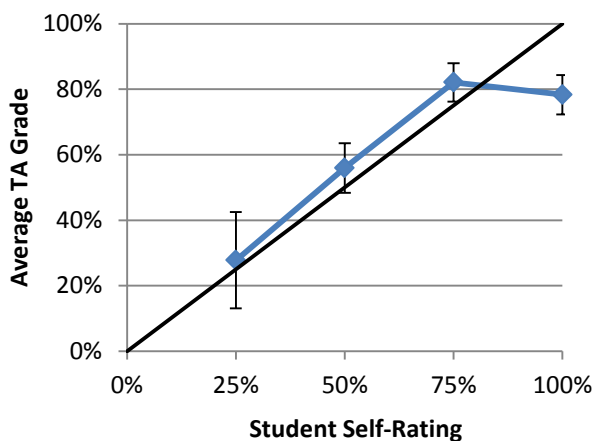
4.6(b) Spectroscopy, Molecular-Level Model:



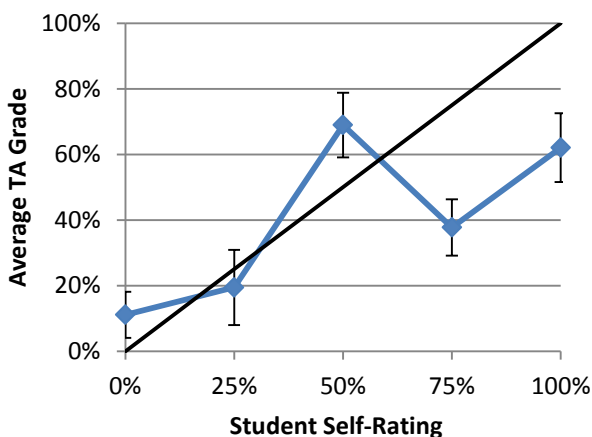
4.6(c) Spectroscopy, Changes to Macroscopic:



4.6(d) Spectroscopy, Changes to Molecular:



4.6(e) Spectroscopy, Use of Evidence:



4.6(f) Spectroscopy, combined:

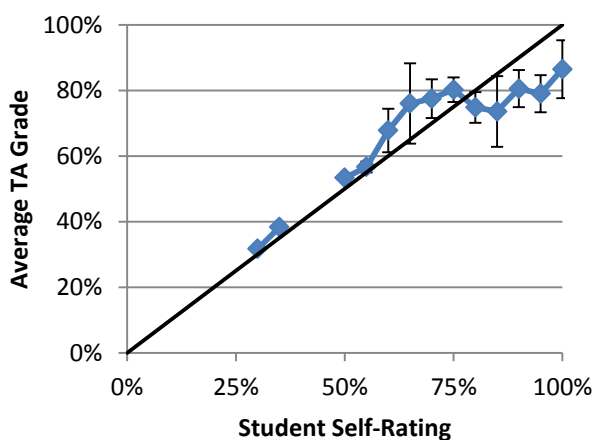
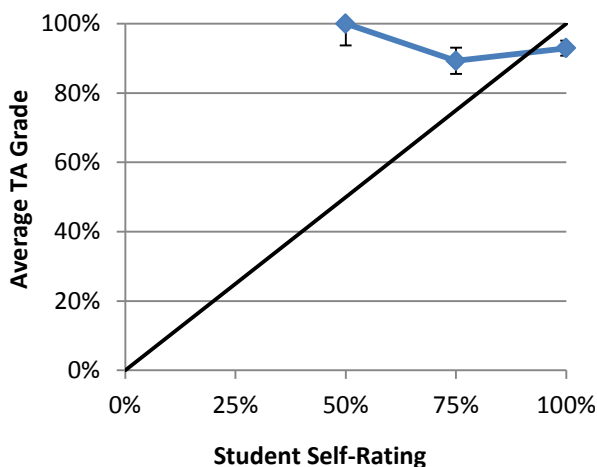
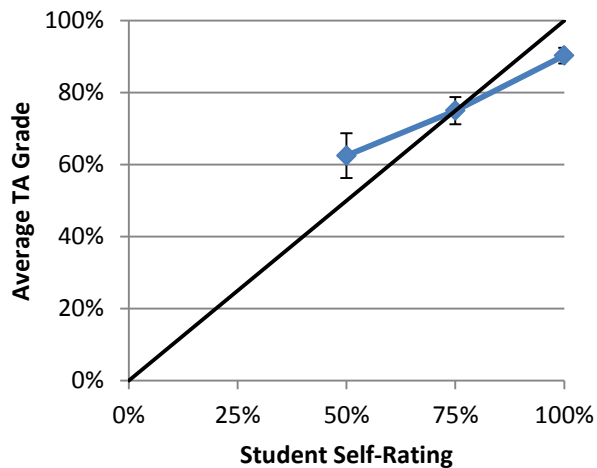


Figure 4.6. Calibration curves for individual model aspects in the Spectroscopy module. The black diagonal line is perfect calibration, as a visual guide; error bars are standard error of the mean. Error bars are not included when a data point represents only one student rating.

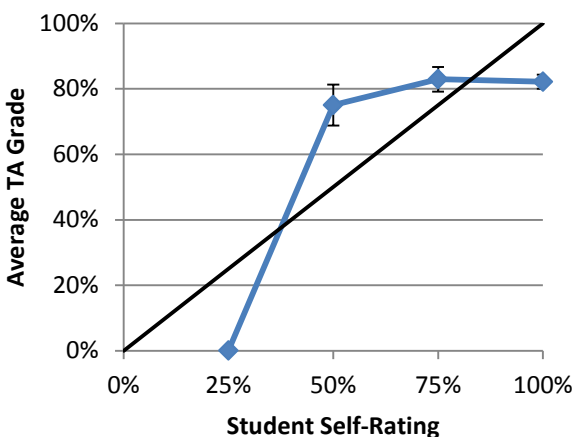
4.7(a) Dissolution, Macroscopic Model:



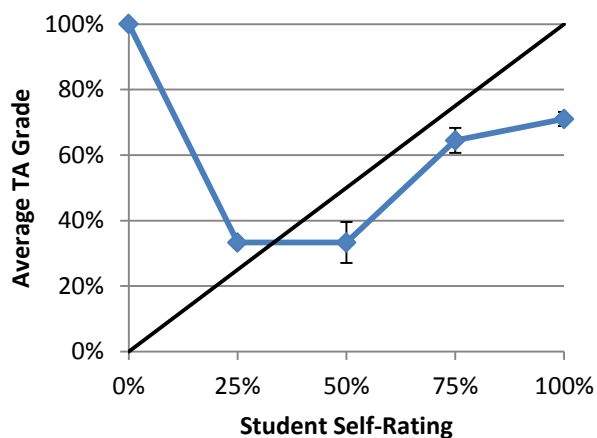
4.7(b) Dissolution, Molecular-Level Model:



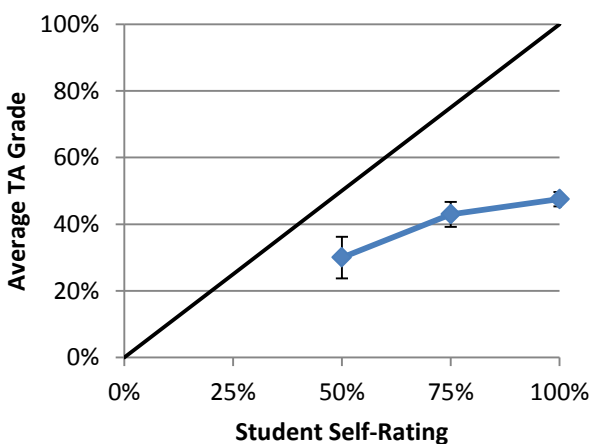
4.7(c) Dissolution, Changes to Macroscopic:



4.7(d) Dissolution, Changes to Molecular:



4.7(e) Dissolution, Use of Evidence:



4.7(f) Dissolution, combined:

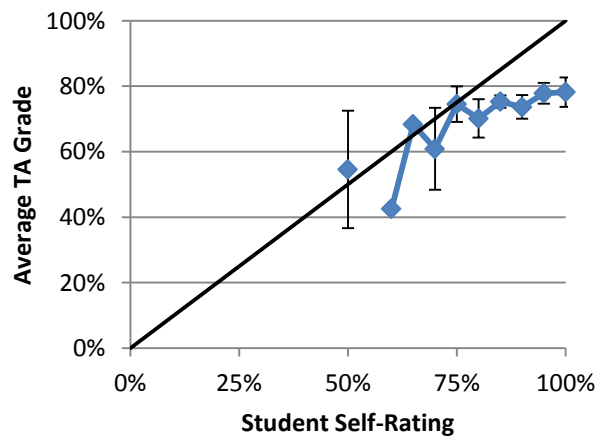
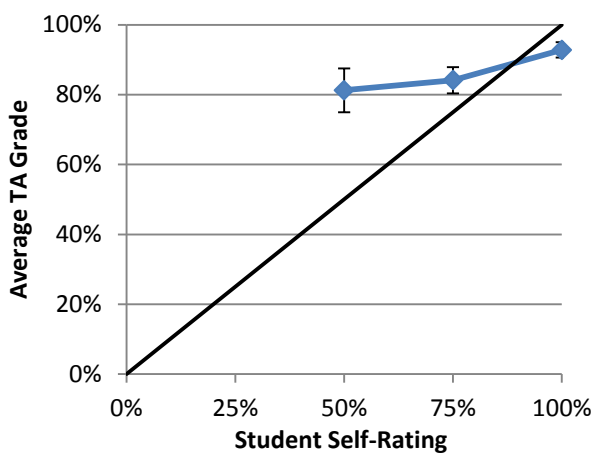
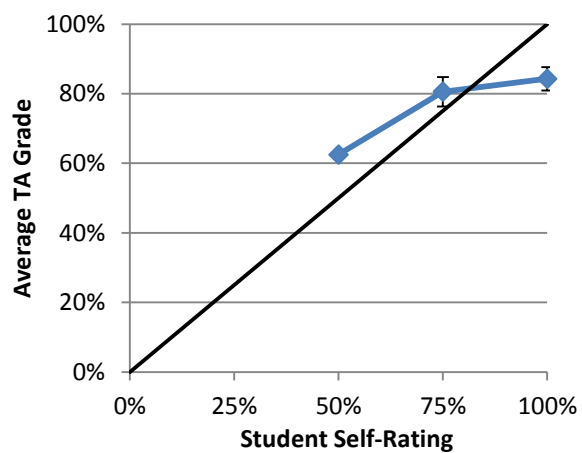


Figure 4.7. Calibration curves for individual model aspects in the Dissolution module. The black diagonal line is perfect calibration, as a visual guide; error bars are standard error of the mean. Error bars are not included when a data point represents only one student rating.

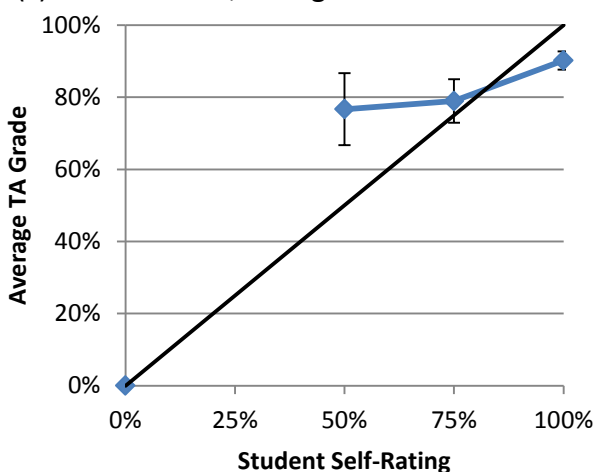
4.8(a) Thermochem, Macroscopic Model:



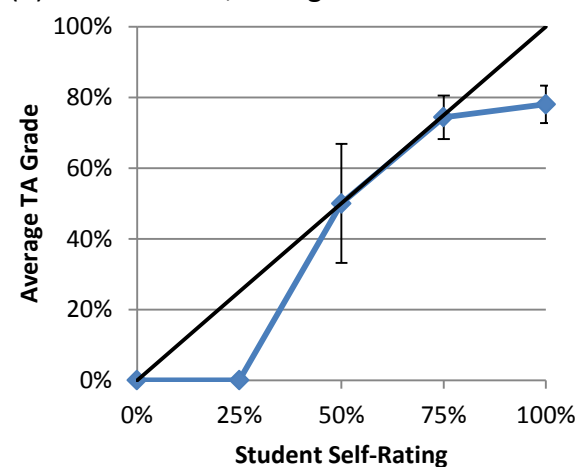
4.8(b) Thermochem, Molecular-Level Model:



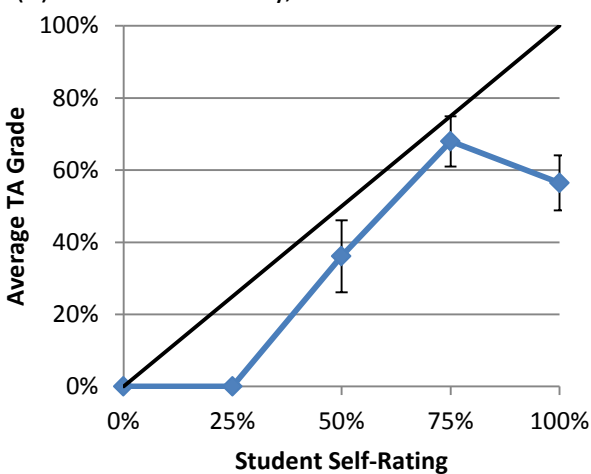
4.8(c) Thermochem, Changes to Macro:



4.8(d) Thermochem, Changes to Molecular:



4.8(e) Thermochemistry, Use of Evidence:



4.8(f) Thermochemistry, combined:

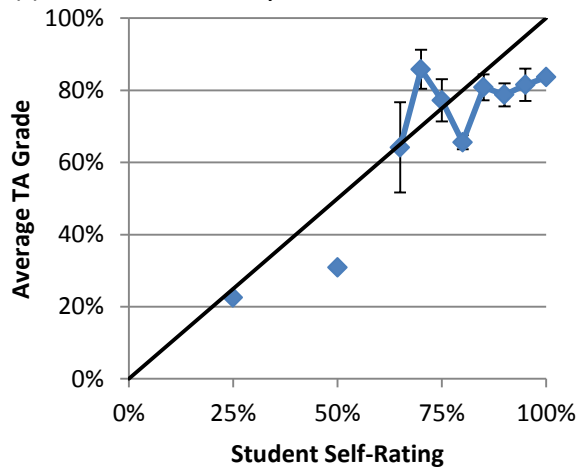


Figure 4.8. Calibration curves for individual model aspects in the Thermochemistry module. The black diagonal line is perfect calibration, as a visual guide; error bars are standard error of the mean. Error bars are not included when a data point represents only one student rating.

Despite the substantial differences between predicted and actual scores described previously, in these charts, it becomes apparent that there are a fair number of instances of good calibration. In fact, for every model aspect in every model, evaluations within 12.5% of the TA grade are the most common or are tied for the most common “bin” of student accuracy; in some cases, more than 50% of student evaluations (most often for presentation of macroscopic and molecular-level models) are accurate to within 12%. A skew towards overestimation of grades is, however, particularly apparent for the three more difficult model aspects – discussion of changes to the macroscopic and molecular-level models and use of evidence.

4.9 a) Student calibration by model aspect, Spectroscopy:

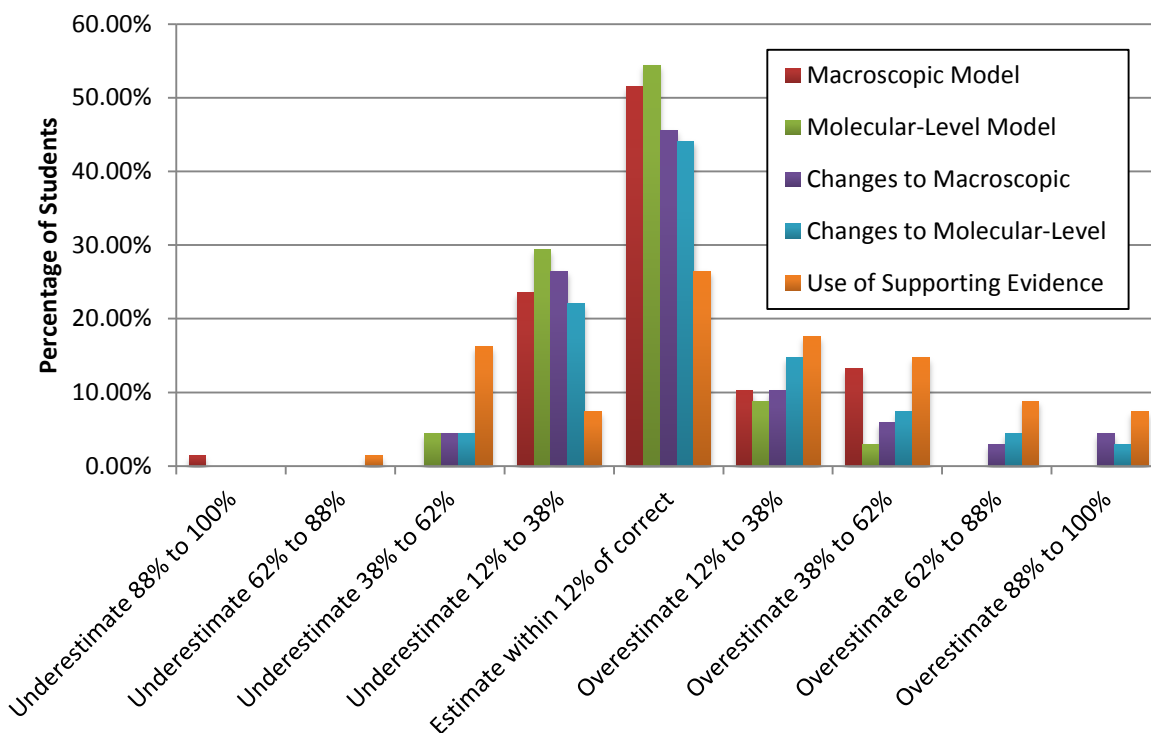
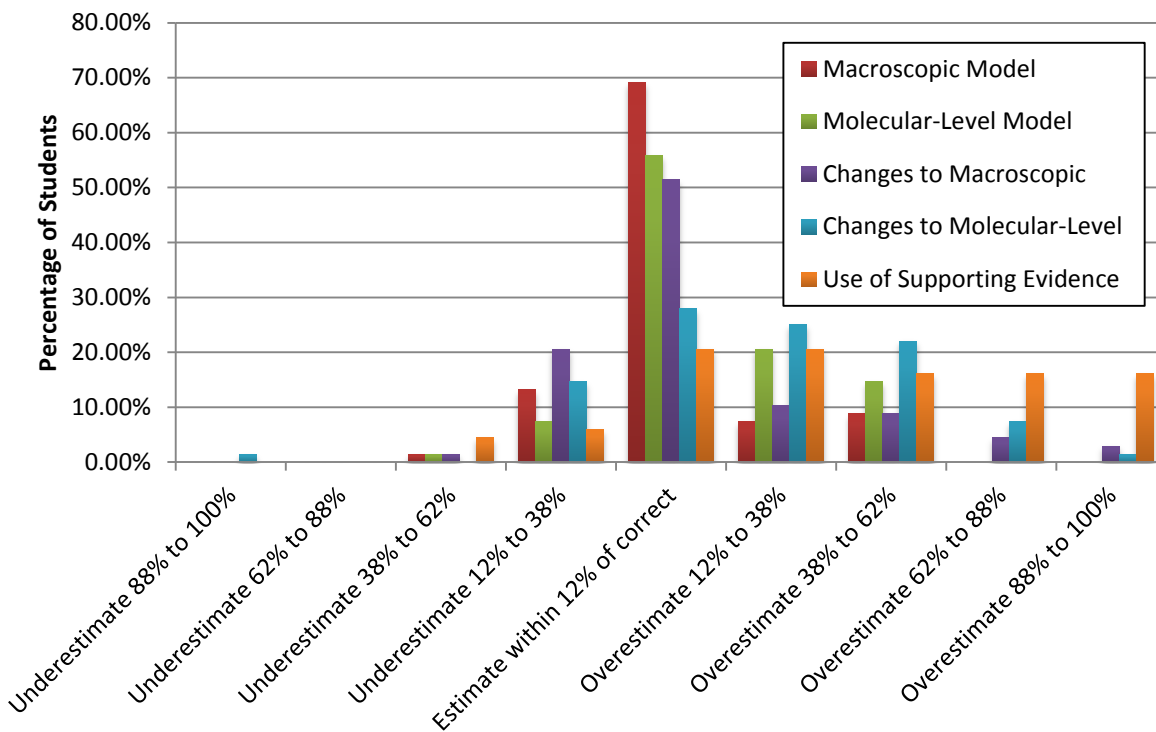
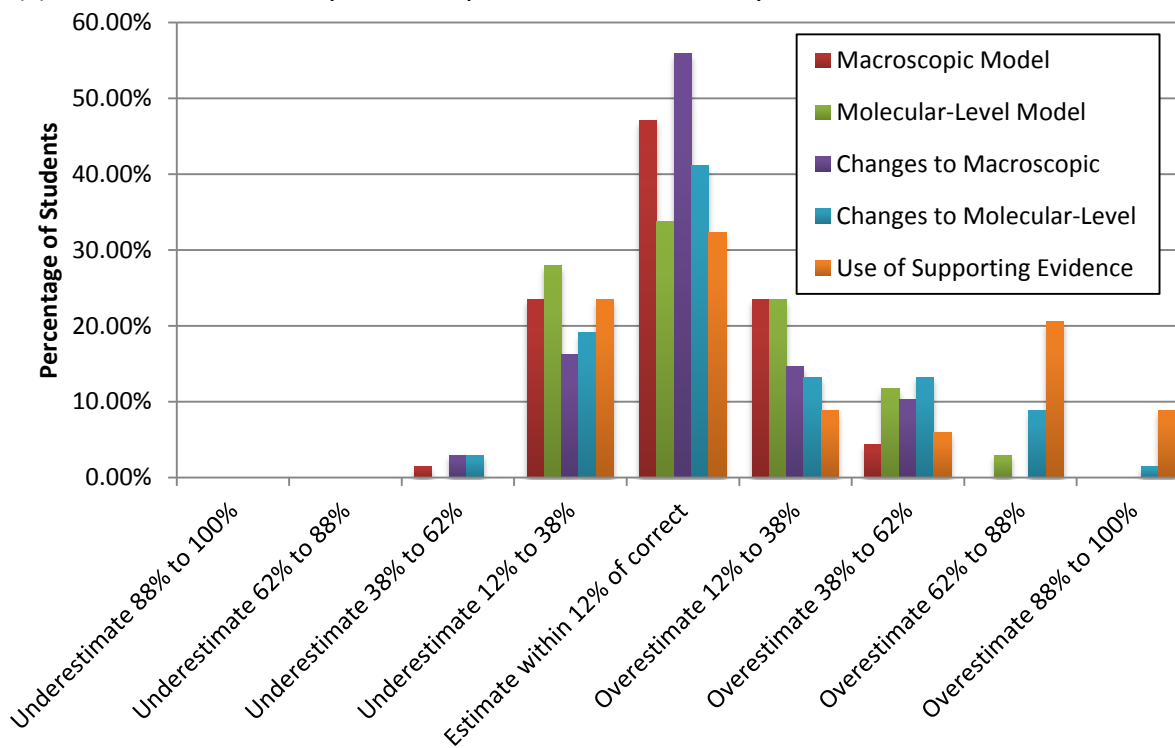


Figure 4.9(a)-(c). [Above and following page.] Distribution of students’ self-assessments relative to TA grades for individual model aspects in each laboratory module analyzed.

4.9 (b) Student calibration by model aspect, Dissolution:



4.9 (c) Student calibration by model aspect, Thermochemistry:



relative to TA grades for individual model aspects in each laboratory module analyzed.

What relationships exist between calibration and students' performance?

Prior research has shown in multiple contexts that lower-performing individuals are also more overconfident in their performance, on tasks including multiple-choice exams^{51, 58}, recall of memorized items,⁶⁹ and reading comprehension.⁷⁰ To examine that relationship in this study, in the more complex writing tasks that students carried out in their models and evaluated using InSPRE, Table 4.8 reports combined calibration values (the difference between students' self-ratings and their TA's grade, averaged across all 5 model aspects) and their relationships with the three judgments of students' performance.

Table 4.8. Correlations between combined calibration scores (derived from self-ratings and TA ratings) and combined ratings from each of the three ratings sources (student self-ratings, TA grade, and expert ratings)

	Spectroscopy	Dissolution	Thermochemistry
Self-Rating vs. Calibration	+0.369, $p=0.002$	+0.361, $p=0.002$	+0.268, $p=0.027$
TA Rating vs. Calibration	-0.690, $p<0.001$	-0.744, $p<0.001$	-0.776, $p<0.001$
Expert Rating vs. Calibration	-0.150, $p=0.223$	-0.336, $p=0.005$	-0.275, $p=0.023$

The positive correlation between student self-rating and calibration reflects the fact that students who rate themselves higher actually are, on average, less accurate in terms of their ratings. This can be seen in the calibration plots, as, in many cases, deviation from the line of perfect calibration is more pronounced for those students who rate their performance at 100% than for those who rate themselves at 75%.

However, the correlations between calibration and TA grades, reported in Table 4.9, broken down both by model aspect and by instructor are much stronger (correlation coefficients ranging from -0.690 to -0.776) than the correlations between calibration and student self-ratings (correlation coefficients ranging from +0.268 to +0.369). Moreover, these correlations are negative, reflecting that smaller (more accurate or underestimated) calibration values are associated with higher TA grades. Intriguingly, these correlations are also stronger than those between student self-ratings and TA ratings – a lower calibration score (tending toward more accurate estimation or underestimation) is a better predictor of higher grade than higher self-rating. These statistically-significant correlations persist for all 15 individual model aspects, as well as for combined scores on each model broken apart by TA (except for TA 1's students in the thermochemistry module). This appears to offer strong evidence that higher-performing students evaluate themselves more accurately – or, rarely, as seen in the histograms from Figure 4.9, underestimate their performance.

However, because calibration is defined as the difference between actual and predicted scores, one should use caution when interpreting the correlation between that calibration score and one of the factors that is used to calculate it.

A more conservative test of whether improved model quality is actually associated with improved calibration would be to use an independent measure as the correlate with calibration. Fortunately, the expert ratings are available as a measure of model quality that does not factor into the calculation of calibration and was consistently applied across instructors. Correlations between calibration (with respect to TA grades) and model quality as judged by expert ratings were also calculated (Table

4.10), and were found to be statistically-significant for both the Dissolution and Thermochemistry lab modules.

Table 4.9. Correlations between TA grades and calibration with respect to TA grades broken down by model aspect (top section) and instructor (bottom section).

	Spectroscopy	Dissolution	Thermochemistry
Combined Ratings	-0.690, $p < 0.001$	-0.744, $p < 0.001$	-0.776, $p < 0.001$
Ratings by Aspect			
Macroscopic Model	-0.752, $p < 0.001$	-0.804, $p < 0.001$	-0.658, $p < 0.001$
Molecular-Level Model	-0.524, $p < 0.001$	-0.828, $p < 0.001$	-0.834, $p < 0.001$
Changes to Macroscopic	-0.754, $p < 0.001$	-0.800, $p < 0.001$	-0.630, $p < 0.001$
Changes to Molecular	-0.716, $p < 0.001$	-0.831, $p < 0.001$	-0.816, $p < 0.001$
Use of Evidence	-0.745, $p < 0.001$	-0.863, $p < 0.001$	-0.898, $p < 0.001$
Ratings by Instructor			
TA 1	-0.730, $p = 0.001$	-0.772, $p < 0.001$	-0.351, $p = 0.183$
TA 2	-0.537, $p = 0.003$	-0.882, $p < 0.001$	-0.791, $p < 0.001$
TA 3	-0.545, $p = 0.006$	-0.649, $p = 0.001$	-0.871, $p < 0.001$

A closer examination broken down by individual model aspect in Spectroscopy finds that no individual model aspect expert quality is significantly correlated with calibration, although a number of model aspects within both Dissolution and Thermochemistry are. Furthermore, the fact that these correlations do not appear in the first laboratory module of the semester, but do appear in later modules, suggests a relationship with the chronological patterns of self-assessments observed earlier.

One of these trends was that students' self-assessments were generally higher in later laboratory modules. However, this is probably not responsible for this trend, as calibration scores did not show any consistent trends across the semester – if any students became more poorly calibrated because of high ratings, others may have

become better calibrated with time. (Changes in individual students' calibration over time were not analyzed in this work, but are an area ripe for further study.)

Table 4.10. Correlations between expert ratings and calibration with respect to TA grades, broken down by model aspect (top section) and instructor (bottom section)

	Spectroscopy	Dissolution	Thermochemistry
Combined Ratings	-0.150, <i>p</i> =0.223	-0.336, <i>p</i>=0.005	-0.275, <i>p</i>=0.023
Ratings by Aspect			
Macroscopic Model	-0.122, <i>p</i> =0.322	-0.414, <i>p</i><0.001	-0.084, <i>p</i> =0.495
Molecular-Level Model	+0.076, <i>p</i> =0.538	-0.273, <i>p</i>=0.024	-0.339, <i>p</i>=0.005
Changes to Macroscopic	+0.070, <i>p</i> =0.568	-0.291, <i>p</i>=0.016	-0.218, <i>p</i> =0.074
Changes to Molecular	-0.114, <i>p</i> =0.354	-0.320, <i>p</i>=0.008	-0.403, <i>p</i>=0.001
Use of Evidence	-0.083, <i>p</i> =0.499	-0.333, <i>p</i>=0.005	-0.343, <i>p</i>=0.004
Ratings by Instructor			
TA 1	+0.311, <i>p</i> =0.242	-0.604, <i>p</i>=0.013	-0.133, <i>p</i> =0.623
TA 2	-0.077, <i>p</i> =0.694	-0.234, <i>p</i> =0.231	-0.189, <i>p</i> =0.335
TA 3	-0.249, <i>p</i> =0.240	-0.268, <i>p</i> =0.205	-0.629, <i>p</i>=0.001

However, students' self-ratings, when differentiated by instructor, mirrored trends in the grades those instructors assigned – but only in later laboratory modules, not in the first. This is also likely to have affected students' calibration with respect to their instructor: when information about their instructor's grading patterns becomes available, some students are likely to attend to this information more than others. Students who are better able to calibrate their own ratings to match those of their instructors are also those who produce higher-quality models.

These data, therefore, suggest that student calibration, or accuracy of self-assessments, is a factor that is related to students' producing models of higher quality –

and, if grading standards are properly aligned, can facilitate student engagement in key thinking processes.

4.7 Discussion

In this chapter, I have identified several factors that are associated with students' self-evaluations – in particular, the aspect of the model that is being evaluated, and the grades delivered previously by the students' instructor. We have seen that students' self-evaluations are general substantially overconfident, but that this overconfidence is not uniformly exhibited across all aspects of models. Even though they recognize that they perform more poorly on this aspect of their models than on other aspects, students are far more overconfident in their use of evidence to support changes to their molecular-level model than on simply presenting their models – although they do recognize the difficulty of this aspect by providing lower self-ratings. Simply put, students generally do not accurately evaluate whether they have appropriately cited evidence to describe whether and how their models have changed, but believe they have adequately done so. They may be less confident about their use of evidence than about their performance on other model aspects, but they are also more overconfident. Previous work has shown that students are sometimes poor at supporting their arguments with evidence⁴²; this research shows that they are similarly poor at evaluating their performance at this critical task.

Students' self-evaluations, however, appear to be influenced in part by the feedback they receive as grades from their instructors. Compared to students of a “stricter” TA, students of a TA who awarded substantially higher grades ended up rating themselves more highly by the end of the term, despite not producing models of higher

quality. Finally, the accuracy of students' self-evaluations – the difference between their self-ratings and their TA's grades – is found to be linked with the quality of their work, with higher-performing students also offering more accurate self-evaluations.

This work extends previous knowledge about students' performance and self-assessments to new types of tasks and situations. Although students' poor performance at supporting scientific explanations with evidence has been documented,⁴² past work has not offered evidence that students' weaknesses at assessing whether they have engaged in this type of thinking may play a part in the quality of their work.

Similarly, this work offers evidence that students' metacognitive monitoring judgments and performance are linked not merely for simple tasks such as recall or general knowledge questions, but also for cognitively complex and more subjective tasks, such as those involved in the writing of MORE models. The expectation that results similar to these would be found informed the design of the portion of this study described in the next chapter, in which we attempted to explore the role of instructors' written feedback in shaping students' metacognitive judgments and their performance.

Chapter 5: Relationships between Written Instructor Feedback, Model Quality and Self-assessment, and Model Resubmissions

As described in the previous chapter, the feedback that students receive in the form of grades appears to have a significant impact on the way in which they self-evaluate their own performance; however, this is only one element of the feedback that students receive. Although some students undoubtedly receive feedback on their models via in-person discussions with instructors or via conversations with peers, the most comprehensive and consistent source of feedback encountered by students is the written comments left by their instructor.

Having found that students' self-evaluations are associated with their performance – and, particularly, that more accurate self-evaluations are associated with better performance – it is reasonable to ask whether improving the accuracy of self-evaluations might be a means by which students' performance can be improved. In essence, we are interested in whether *metacognitive monitoring* can be improved so that students may exercise more effective *metacognitive control*.

While the conception of the importance of *metacognitive monitoring* underlies much of the research tradition surrounding metacognition in cognitive psychology, it is by no means unique to that discipline. Research in the field of educational psychology has similarly focused on the importance of self-monitoring and self-regulation, and this has been of particular interest in constructing models that explain why and in what circumstances feedback is effective.⁷¹ Much of the research surrounding feedback has

identified characteristics of feedback that contribute to its effectiveness, including timing, source of feedback, and the nature of the task for which feedback is given.^{71b} Less attention has been paid to the content of feedback itself ^{71a} – somewhat understandably, as this is harder to quantify and control, particularly in authentic classroom settings.

However, when researchers studying feedback have postulated mechanisms by which feedback functions, self-regulation lies at the center.⁷² Referring back to descriptions of self-assessment, particularly those of Paris and Winograd ⁷³, Hattie and Timperley note ^{71b}: “There are two major aspects of self-assessment: self-appraisal and self-management. Self-appraisal relates to students’ facility to review and evaluate their abilities,” whereas self-management entails the “regulating of students' ongoing behavior through planning,” and other strategies. These two components of self-assessment – self-appraisal and self-management – parallel the definitions of *monitoring* and *control* employed in the metacognition literature.⁷⁴

This suggests a particular cognitive skill – self-regulatory ability – that the actual written text of feedback may be able to target. Hattie and Timperley ^{71b}, in their review of the feedback literature, identify four common targets of feedback. Briefly, these are (with examples from Hattie & Timperley):

- Feedback about the self, which is frequently affective in nature and provides little task-related information and has little bearing upon student achievement. (“You’re a great student.”)
- feedback about the task, which is directed toward the quality of a student’s performance and provides primarily surface-level information (“You need to include more information about the Treaty of Versailles”);
- feedback about the process, which is directed toward the processes involved in

students' performance and relates to facets of performance such as error-detection and strategy searching. ("This page may make more sense if you use the strategies we talked about earlier.")

- feedback about self-regulation, which is directed towards students constructing their own strategies for assessing and improving their performance. ("You already know the key features of the opening of an argument. Check to see whether you have incorporated them in your first paragraph.")

Developing the capacity for self-regulation has been repeatedly identified as a key goal of effective feedback.⁷⁵ Sadler⁷⁶ argues that a major challenge for learners "is to assimilate the teacher's feedback into their existing knowledge bases so that it can be drawn upon, as needed, in future constructive activity" and that feedback which is merely oriented students towards "telling" students about the quality of their work is of limited effectiveness. While Hattie and Timperley^{71b}, in their review, do not identify studies that have specifically delivered feedback aimed directly at the self-regulatory level, a few such studies have been conducted in the realm of cognitive psychology.

In particular, Glenberg, Sanocki, Epstein, and Morris⁷⁷ and Walczyk and Hall⁵⁰ demonstrated improvement in reading comprehension through inclusion of pretest questions designed to assist students in creating self-generated feedback. If students are assisted in self-generating feedback, then one would expect both students' self-monitoring accuracy (calibration, as described in the previous chapter) and the quality of students' models to be improved.

The central question I sought to examine with these aspects of our study, therefore is: *Does receipt of feedback targeted at students' monitoring of their own*

performance at model-writing improve the accuracy of students' self-assessments and/or the quality of the models they produce?

5.1 Study Design

This study was conducted concurrently with the work described in the previous chapter; therefore, the study population and laboratory curriculum are the same as those described in Chapter 4. This enables us to use the same student self-assessments, collected using the InSPRE activity, and the same expert model quality ratings.

Before instructors graded students' first model assignments, the student population from the five study sections was randomly divided into two groups. To ensure that the two groups were comparable, the potential groups were checked in order to verify that there were not statistically-significant differences between them with respect to gender ratios, scores on the American Chemical Society's Examinations Institute 1998 Toledo Chemistry Placement Examination (administered at the start of the term), or performance on the first activities of the semester (a worksheet and quiz on laboratory safety). As discussed in Section 5.2, however, because not all students completed all of the assignments and activities analyzed as part of this study, these comparisons do not necessarily hold for the final populations of students (average scores on the Toledo Exam are reported in Table 5.4).

Instructors were provided with instructions on delivering feedback for students; these instructions are reproduced in Figure 5.1. In brief, instructors were told to provide one group of students (the "direct feedback" condition) with feedback that informed them about deficiencies in their models and identified specific areas where students should have performed differently to have a more complete or higher-quality model. For

the other group of students (the “reflective feedback” condition), instructors were asked to provide feedback that did not directly inform students of deficiencies in their models, but asked students to search for information within their models that would enable them to identify those issues – preferably using questions or statements that would be general enough to be applicable for other refined model assignments as well. The instructions also offered examples of both “direct” and “reflective” feedback that might apply to particular situations. The author also made himself available to other instructors in the study to offer further guidance about the types of feedback intended in these two conditions.

Analyses of feedback provided by experienced MORE instructors in a semester prior to this study (including TA 2 and TA 3, but also including other instructors) revealed that a relatively small fraction of written feedback –typically less than 25% -- pertained to discussion of model changes and supporting evidence – key aspects of students’ written work. Therefore, instructors in this study were directed to target at least 50% of the feedback instances on each refined model towards these elements of student performance, and to make at least two written comments on these aspects for each student’s model. By providing quantitative guidelines for the amount and quantity of feedback, I hoped to eliminate discrepancies that have been observed in the literature between instructors’ personal impressions about their written feedback and the actual feedback that was delivered. ⁷⁸

Feedback Instructions for Fall 2008 Research Sections

For each student paper:

- Write feedback for each instance where a students' model is marked down as incomplete. Do not mark associated point deductions alongside this feedback.
- Write at least two feedback comments on each refined model regarding students' description of model changes.
- Write at least two feedback comments on each refined model regarding students' use of experimental evidence.
- Check that at least half of the feedback you provide on refined models is related to discussion of model changes and use of evidence.

For each feedback instance addressing a deficiency ("incomplete") in a model, your feedback should be either reflective (monitoring-oriented) or direct (task-oriented) depending on which research group (see following pages) the student is assigned to.

Reflective (monitoring-oriented) Feedback:

- 1) Should prompt students to search for information, particularly information about the goal we intend the student to achieve.
- 2) Should direct students to contrast their performance with a benchmark level of performance.
- 3) Should be generally relevant to any model assignment, not particular to a student's deficiencies in a particular model. (Many reflective feedback statements may be similar).

Direct (task-related) feedback:

- 4) Should provide information directly to students as to what they need to improve in their models.
- 5) Should be particular to that student's performance on that particular model assignment.

Direct feedback

Include macroscopic observations with multiple senses.

What happens when sugar is added to water?

Be more specific about changes to your molecular-level model.

What evidence supports this change?

Reflective feedback

Does this include all the types of observations you might make?

Review the assignment to make sure you've described every chemical system.

Have you described the changes to every aspect of your model?

Did you describe what evidence supports your new model?

Figure 5.1. Feedback instructions provided to instructors during the fall 2008 semester.

5.2 Analysis of Written Feedback Delivered

It is apparent from even a cursory examination of students' graded models that the extent and types of feedback differed markedly between instructors. TA's reports during and after the term corroborated this impression. To characterize the feedback delivered – and particularly the degree to which TAs faithfully implemented directions for the “direct” and “reflective” feedback conditions – feedback statements were analyzed.

After the conclusion of the semester, I transcribed the feedback comments written by the TAs on the papers of consenting students. This totaled over 9,000 feedback statements. Narrowing the data set to the 68 students considered in these analyses yielded 6,895 feedback statements delivered to students. TA's feedback addressed a range of topics, from the key aspects of students' models discussed in depth in this study – presentation of models, discussion of model changes, use of supporting evidence – to recording of data and observations, in-class calculations, and formatting of laboratory manual pages. During transcription, the section of the model or lab report that each feedback statement related to was also recorded. In some cases, statements pertained to more than one aspect. The aspects of students' assignments to which feedback statements related are identified in Table 5.1, along with the number of feedback comments on each particular aspect of the model assignment per student over the course of the entire semester. The set of assignments analyzed for feedback includes nine refined models and four initial models graded and returned to students before they submitted the final refined model of the semester.

Table 5.1. Number of feedback statements per student delivered on model assignments over the semester, by section of model assignment and instructor. Bolded model assignment sections are those focused on in this study. Italicized sections fall outside the “model” portion of the laboratory reports.

Model Section or Type of Feedback	TA 1	TA 2	TA 3
Macroscopic Model	14.3	3.5	9.9
Molecular-Level Model	50.4	3.3	19.5
Changes to Macroscopic Model	16.0	4.2	10.3
Changes to Molecular-Level Model	34.5	4.9	11.9
Evidence for Molecular-Level Changes	18.3	7.6	13.4
Generalization	24.3	2.4	10.6
Use of Evidence to Support Generalization	14.3	3.5	9.5
Description of Proposed New Experiment	12.7	3.5	10.9
<i>Reporting of Data and Observations</i>	24.8	4.6	11.9
<i>Post-Lab Questions</i>	5.2	1.9	2.3
<i>Procedural or Formatting Matters Not Related to Model Content</i>	0.4	0.6	0.2

The most pronounced contrast between instructors is that the quantity of feedback delivered by each TA differed widely. TA 1 provided an average of 217.1 feedback statements to each student over the course of the semester, while TA 2 provided an average of 40.0 feedback statements per student. TA 3’s level of feedback fell between the two, providing an average of 110.3 feedback comments per student over the entire semester. Interestingly, this trend parallels that of the grades assigned by instructors – TA 1, who assigned the lowest average grades, provided the most feedback, while TA 3, who assigned the highest grades, provided the least feedback.

Despite these differences, the fraction of feedback comments delivered by each TA on important aspects of the assignment was surprisingly consistent. Each TA delivered between 82 and 87% of their feedback on the model portions of the laboratory

assignments, and between 58% and 62% of their feedback on the five major aspects discussed in this dissertation. Because these five aspects are the primary focus of students' self-evaluations and model quality coding, they are also the focus of the more detailed analysis of feedback that follows.

To investigate the degree to which TAs delivered feedback aligned with the directions they received and to which students in each feedback condition received contrasting types of feedback, written feedback statements were coded for two attributes associated with reflective feedback:

- “search” statements - statements that ask students to examine their models or the assignment for the answer; questions that, to answer, would require the responder to read the model if they weren't familiar with it (“Have you identified what evidence supports your new ideas about dye molecules?”)
- “general” statements – statements or questions which, ignoring the context of the model they are found in, would be a reasonable general idea or advice to consider when writing any model (“Always support molecular-level changes with reference to evidence.”)

(Note that, in the examples given above, the “search” statement is not necessarily general – it refers to specific aspects of a model – and the “general” statement does not necessarily call for a search – it prescribes content to include rather than asking the student to examine their model. The statement “Have you supported your molecular-level changes with evidence?” would be classified as both “search” and “general.”)

Tables 5.2 and 5.3 describe the degree to which the feedback statements delivered fall into these categories, broken down by both instructor and feedback condition. Because I was specifically examining the effect of feedback targeted towards self-

regulation, hallmarks of “direct” feedback were not coded for expressly. Both the number of statements of each type (“search” or “general”) and the percentage of total feedback statements received by students within that group out of the total set of feedback statements are included. Both the quantity of reflective feedback statements and the proportion of feedback that is of that type may be relevant factors, as it is plausible that students may need to see reflective feedback frequently and consistently in order for questions similar to those asked in reflective feedback to become a consistent part of students’ own self-critiquing processes. These two measures sometimes diverge – for instance, according to Table 5.2, TA 1 and TA 2 both delivered a comparable number of “search” feedback statements on the molecular-level aspect to the (1.4 vs. 1.7). However, for TA 2, this amounts of 53% of the feedback delivered to these students, whereas, for TA 1, who delivered much more feedback in total (as described in Table 5.1), this was only 3% of the feedback delivered to those students.

For every pairing between direct and reflective feedback groups, and students in the reflective feedback group received a greater proportion of search statements ($p < 0.002$ for comparison overall proportion of “search” feedback delivered across all model aspects). However, while for TA 2 and TA 3, this difference is pronounced— less than 10% of comments written to the direct feedback group on each aspect were search statements and more than 50% of comments written to the reflective feedback group were “search” statements – this contrast is much less pronounced for TA 1, where for no aspect was more than 9% of the feedback delivered as “search” or “general” statements. Furthermore, TA 1 delivered “general” feedback to direct feedback students roughly as often (as measured by proportion of feedback delivered) as to reflective feedback students (a non-significant difference between groups; $p = 0.284$). This instructor also

reported not consistently applying the feedback condition after the first few models of the semester.

TA 2 was best at consistently delivering reflective feedback statements on the portions of models relating to model changes, giving “search” feedback over 85% of the time and “general” feedback over 65% of the time. This contrasts with their feedback concerning presentation of macroscopic and molecular-level models, which was only about 50% “search” and 10% “general” within the reflective feedback group.

Table 5.2. Average number of “search” statements received per student (and, in parentheses, fraction of total feedback statements on that model aspect that are “search” statements), broken down by TA and feedback groups and by the portion of the model which the statements apply to.

		TA 1, Direct Condition	TA 1, Reflective Condition	TA 2, Direct Condition	TA 2, Reflective Condition	TA 3, Direct Condition	TA 3, Reflective Condition
Number of “Search” Statements Pertaining to Each Aspect	Macroscopic Model	0.1 (1%)	0.9 (7%)	0.1 (4%)	2.0 (51%)	0.3 (2%)	6.3 (78%)
	Molecular- Level Model	0.3 (0%)	1.4 (3%)	0.2 (6%)	1.7 (53%)	0.3 (1%)	11.5 (72%)
(Percentage of “Search” Statements Out of Total Feedback Statements Pertaining to that Aspect)	Changes to Macro Model	0.1 (1%)	1.0 (6%)	0.1 (7%)	5.1 (89%)	0.3 (2%)	7.2 (80%)
	Changes to Molec Model	0.0 (0%)	0.7 (2%)	0.2 (9%)	5.5 (87%)	0.3 (2%)	8.2 (81%)
	Evidence Supporting Changes	0.0 (0%)	0.7 (3%)	0.3 (5%)	5.4 (63%)	0.2 (1%)	7.6 (62%)

Table 5.3. Average number of “general” statements received per student (and, in parentheses, fraction of total feedback statements on that model aspect which that are “general” statements), broken down by TA and feedback groups and by the portion of the model the statements apply to.

		TA 1, Direct Condition	TA 1, Reflective Condition	TA 2, Direct Condition	TA 2, Reflective Condition	TA 3, Direct Condition	TA 3, Reflective Condition
Number of “General” Statements Pertaining to Each Aspect	Macroscopic Model	0.7 (4%)	1.1 (9%)	0.1 (4%)	0.4 (11%)	0.9 (8%)	2.5 (31%)
	Molecular- Level Model	3.3 (6%)	3.1 (7%)	0.1 (3%)	0.3 (11%)	2.3 (10%)	5.8 (36%)
(Percentage of “General” Statements Out of Total Feedback Statements Pertaining to that Aspect)	Changes to Macro Model	1.7 (11%)	0.9 (6%)	0.2 (13%)	4.1 (73%)	1.4 (12%)	3.8 (43%)
	Changes to Molec Model	2.6 (7%)	1.8 (6%)	0.1 (4%)	4.2 (66%)	1.7 (12%)	4.3 (43%)
	Evidence for Molec Changes	0.4 (3%)	0.8 (4%)	0.0 (4%)	2.8 (33%)	2.3 (16%)	4.0 (33%)

TA 3’s implementation of the feedback directions was the most consistent across all model categories, consistently giving between 62% and 81% “search” feedback and between 31% and 43% “general” feedback to the reflective feedback group, while making statements to the direct feedback group that were classified as “search” statements less than 2% of the time and “general” statements only between 8 and 16% of the time (overall proportion of both “search” and “general” feedback was significantly different between groups; $p < 0.001$ for both comparisons)

TA 2 and TA 3 can be considered to have faithfully implemented the directions concerning the type of feedback to deliver throughout the semester, while TA 1, who offered the most feedback, did not consistently produce a contrast in the feedback delivered to the two groups. Although TA 2 delivered less feedback overall, she delivered

the strongest contrast in feedback between direct and reflective feedback groups on the model aspects pertaining to discussion of changes. However, she was not as consistent in delivering contrasting feedback to the two groups for other model aspects. TA 3 delivered pronounced and consistent contrasts between the two feedback conditions across all model aspects, delivering the greatest quantity of reflective feedback to students in that condition.

5.3 Comparisons Between Feedback Groups

One logical way to examine the effects of feedback on students is to directly contrast the performance of students in the reflective feedback condition with those in the direct feedback condition on the various measures discussed in Chapter 4: each of the three measures of model quality (student self-ratings, TA grades, and instructor ratings) as well as calibration scores (specifically, calibration of students' self-evaluations with respect to TA grades, as discussed in Section 4.5).

Comparisons between feedback conditions can be made both for all students globally or for each individual instructor. Because it has been seen in the previous section that each of the instructors delivered written feedback of substantially differing quantities and types, each instructor's cohort of students will be considered separately in analysis by feedback condition despite the loss of statistical power due to smaller populations in each comparison. TA 2 and 3's populations could be arguably be combined in these comparisons due to the substantial contrasts between feedback delivered in the two conditions, however, the substantially different quantities of feedback delivered by those two instructors may impact students' responses to that feedback, and so they are still considered separately.

First, the two groups in each instructor's student population were examined to determine if there were significant differences in their performance on the pretest that might indicate differences in the narrowed student populations for each feedback condition. Although, as described previously, assignment of students to feedback conditions was done in a manner that attempted to equalize average pretest scores in each group of consenting students, both the fact that students were added to the study population when they consented later in the semester and that students were dropped from the study population because they did not complete some of the activities analyzed mean that the matching of pretest averages does not necessarily persist in the final study population. Table 5.4 reports average scores for the mathematics and chemistry-general subsections of the pretest; the differences between each instructor's "direct" and "reflective" feedback groups' averages are not statistically significant. (Although this does not mean that the two groups' performances are statistically equivalent, possible impacts of individual differences in pre-term measures such as Toledo exam score are discussed in Section 5.5.)

Table 5.4. Population sizes and average scores on the Mathematics (“Math”) and General Chemistry Knowledge (“Chem”) subsets of the ACS Toledo general chemistry placement exam.

	Direct Feedback		Reflective Feedback	
TA 1	<i>n</i> =7		<i>n</i> =9	
	Math=15.6/20	Chem=10.7/20	Math=16.9/20	Chem=11.7/20
TA 2	<i>n</i> =10		<i>n</i> =18	
	Math=14.7/20	Chem=10.2/20	Math=16.5/20	Chem=11.8/20
TA 3	<i>n</i> =12		<i>n</i> =12	
	Math=16.9/20	Chem=11.4/20	Math=16.9/20	Chem=10.9/20

How does students’ performance at model-writing vary between feedback conditions?

Average scores on both external ratings – those generated from the expert model quality coding scheme and those provided by the TA in the form of grades – are presented in Figure 5.2, broken down by individual instructors’ student populations, model assignment, and feedback group. If the feedback condition had a significant effect on model quality, one might expect that model quality ratings might be roughly comparable at the start of the term (in the Spectroscopy module) and that differences would appear by the end of the term (in the Thermochemistry module).

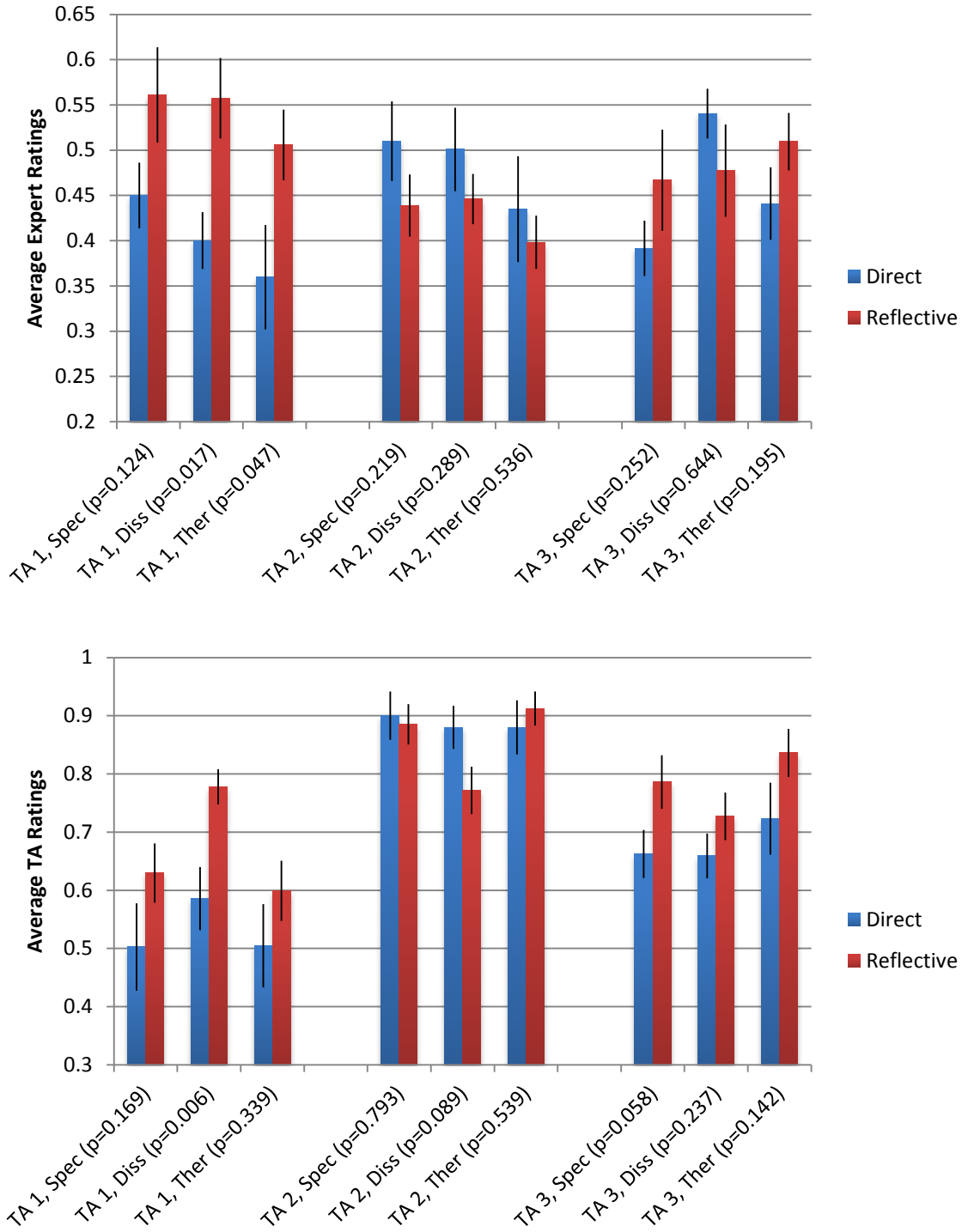


Figure 5.2. Expert (a) and TA (b) ratings of student model quality for each TA’s students and each model analyzed, broken down between “direct” and “reflective” feedback conditions. Error bars are standard error of the mean. Reported p-values are two-tailed independent samples t-tests for model quality ratings between the two feedback conditions.

Only three of the comparisons between groups on any model assignments show statistically-significant differences between the two groups in terms of performance. All of these contrasts occur within TA 1's student population, and all favor the reflective feedback group. Students in the reflective feedback condition produced higher quality models as assessed by the expert rater ($p=0.047$, independent samples t-test) within the thermochemistry module, and by both their instructor ($p=0.006$) and the expert rater ($p=0.017$) in the Dissolution module. However, as described in Section 5.2 above, TA 1 showed the poorest contrast between feedback delivered to the direct and reflective feedback groups; additionally, although the difference is not significant, average TA and expert ratings for TA 1's reflective feedback students were already higher than those for direct feedback students in the first model of the semester (Spectroscopy). It is therefore unlikely that differences in feedback are responsible for any divergence in student performance.

By contrast, TA 2 and TA 3 provided markedly different feedback to students in each of the two feedback conditions, but there are no statistically significant differences between the two groups' performance at any point during the semester – as measured both by expert ratings and by TA grades – in either TA's student populations. This offers further evidence that the feedback variations conducted in this study did not have substantial impacts on students' production of models within laboratory modules later in the semester.

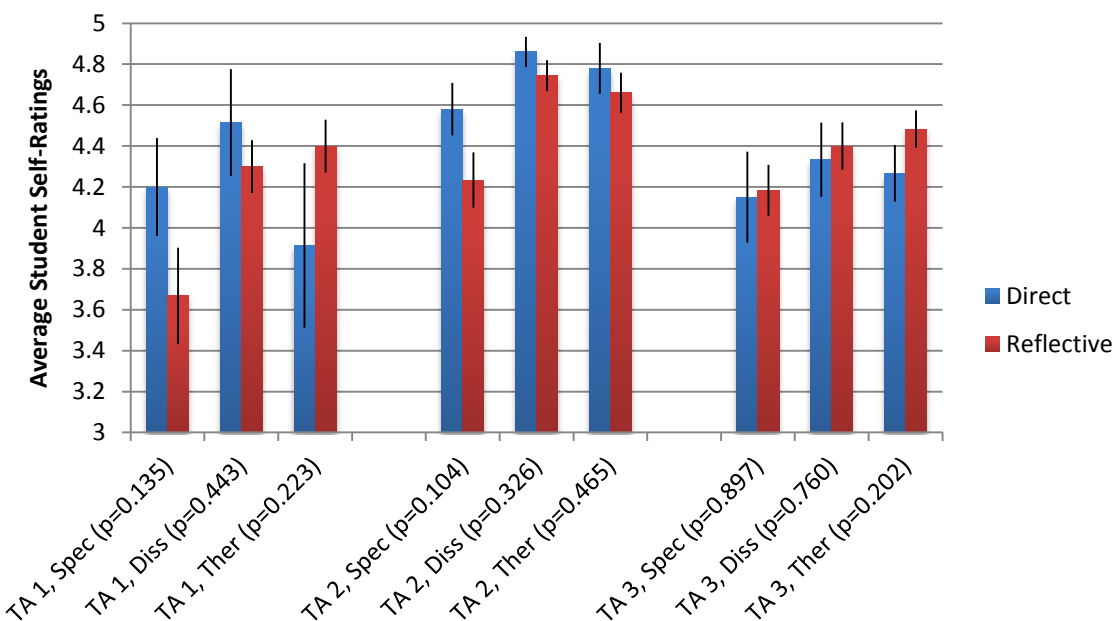


Figure 5.3. Average self-assessments for students, broken down by instructor, model assignment, and feedback group. Error bars are standard error of the mean. The p-values reported in x-axis labels are independent samples t-tests comparing the above direct and reflective feedback groups.

How does feedback condition affect students' self-assessments?

Average calibration, the difference between students' self-assessments and performance, is charted in Figure 5.4 as a function of TA, model, and feedback condition, and shows no consistent patterns in students' self-assessments. There are statistically significant differences between the direct and reflective feedback groups for only for the Spectroscopy and Dissolution models within TA 1's student population. In these cases, the reflective feedback group shows significantly more accurate calibration than the direct feedback group. This is consistent with the self-ratings and TA ratings reported above. In these cases, students in the "direct feedback" condition rated

themselves more highly than students in the “reflective feedback” condition, while producing lower-quality models. However, it is not likely that the feedback manipulation is the causes of these differences, as TA 1 did not deliver substantially different feedback to the two groups throughout the semester. Even within the earlier weeks of the term, these differences in feedback were small compared to those seen between TA 2 and TA 3’s direct and reflective feedback groups.

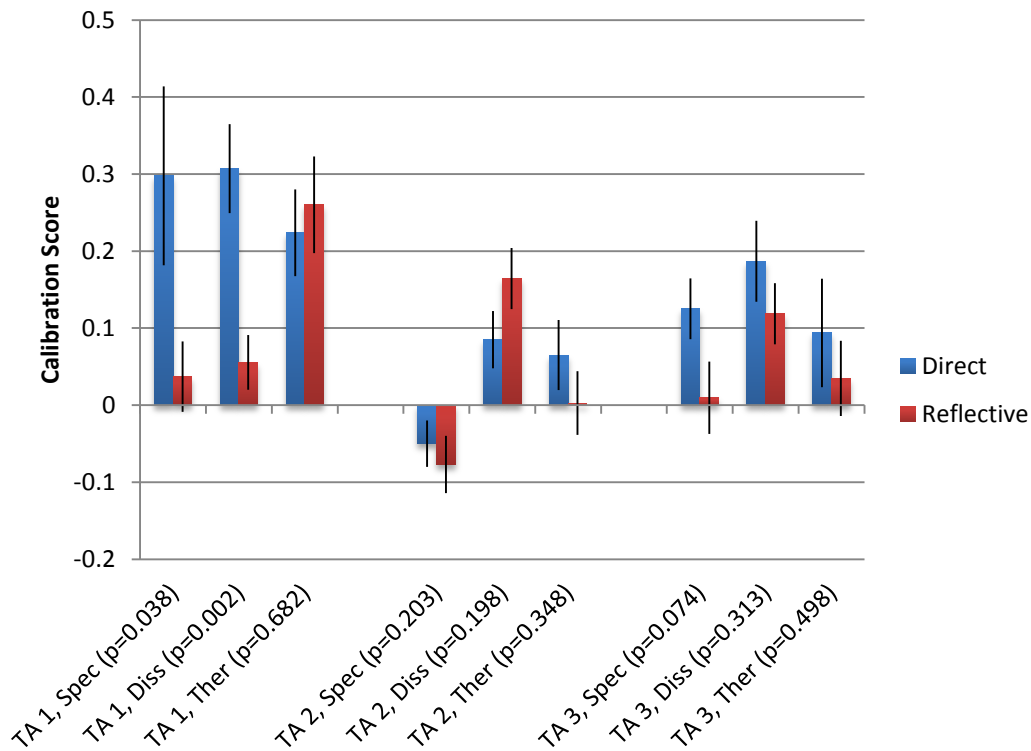


Figure 5.4. Average calibration (TA rating minus student self-rating) for students, broken down by instructor, model assignment, and feedback group. Error bars are standard error of the mean. The p- values reported in x-axis labels are independent samples t-tests comparing the above direct and reflective feedback groups.

The lack of clear differences in performance between the direct and reflective feedback groups, combined with the differences in feedback delivered by each instructor, suggests that a more fine-grained examination should be made of students' self-assessment, performance, and particularly calibration as a function of the actual feedback received by each individual student. In these analyses, aspects should include not only the particular types of feedback received (direct, reflective, or otherwise), and the particular aspects of the model in question (presentation, discussion of changes, use of evidence, etc.), but also the timing of the feedback. Only the feedback received before a particular self-evaluation should be considered in developing a model of students' self-assessment or calibration.

5.4 Analysis by Feedback Received

The above statistics describing the average feedback delivered by each instructor to students in each feedback condition gloss over the fact that the attributes of feedback received by each individual student vary over a significant range, even within groups. As each individual student can only be expected to respond to the feedback they directly received, and not to the "average" or "typical" feedback delivered to a student in the same condition, it is important to characterize the particular feedback received by each individual student. This characterization can then enable the identification of subgroups in which a clearer contrast exists between direct and reflective feedback conditions, or enable the identification of feedback-related variables that may contribute to a model of student performance. (An alternative approach, based on examining correlations between characteristics of feedback received and end-of-semester performance measures, is described in Section 5.5) Additionally, the effects of feedback on particular

model aspects should be closely examined. Here, I have chosen to focus on two aspects of students' models – discussion of changes to molecular-level models and use of evidence to support those changes – because of both their complexity and their relationship to the key thinking processes associated with transfer.

Among the first questions to answer is whether there is an identifiable population of students who received primarily “reflective” feedback on certain model aspects? To identify the students in which reflective feedback was most concentrated, I first identified feedback statements that were coded as both “search” and “general” statements. Of the 4,004 feedback statements delivered to the study population, only 559 (14.0%) meet those criteria. Of these, 125 pertain to students' discussion of changes to their molecular-level models, and 94 to students' use of evidence to support molecular-level changes. Because nearly 40% of the reflective feedback statements relate to these two aspects of student models, it is these aspects that are examined more closely below.

The figures on the following pages present histograms detailing the number of students who received particular quantities of reflective feedback statements (defined as the number of statements coded as both “search” and “general”) or particular percentages of reflective feedback (out of the total number of feedback statements delivered). Figure 5.5 presents this information for the “discussion of molecular-level changes” aspect of students' models, while Figure 5.6 presents this information for the “use of evidence to support molecular-level changes” aspect. As might be expected from the summary data presented above for each instructor and feedback condition, only students in the reflective feedback groups of TA 2 and TA 3 received greater than 30% reflective feedback or more than 3 reflective feedback statements on that particular

model aspect throughout the semester, although some students within those groups also received less reflective feedback than this. Feedback condition, as previously suggested, is not an ideal identifier of students who received the most reflective feedback.

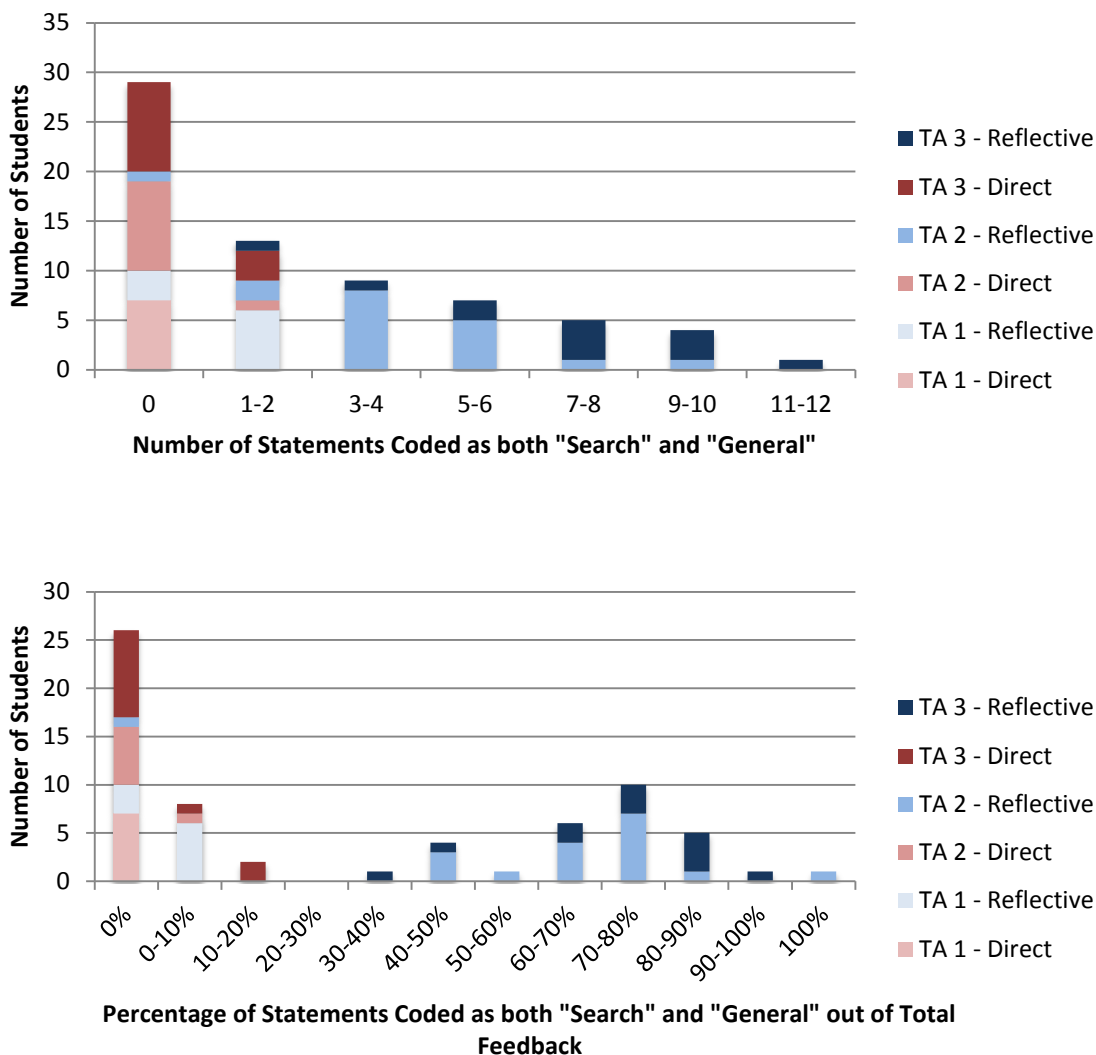


Figure 5.5. Histograms of (a) the total number of students receiving various quantities of reflective feedback statements (those coded as both “search” and “general”) during the semester, (b) the total number of students who received various proportions of reflective feedback statements during the semester. Both histograms refer only to feedback statements pertaining to the “discussion of molecular-level changes” aspect of students’ models.

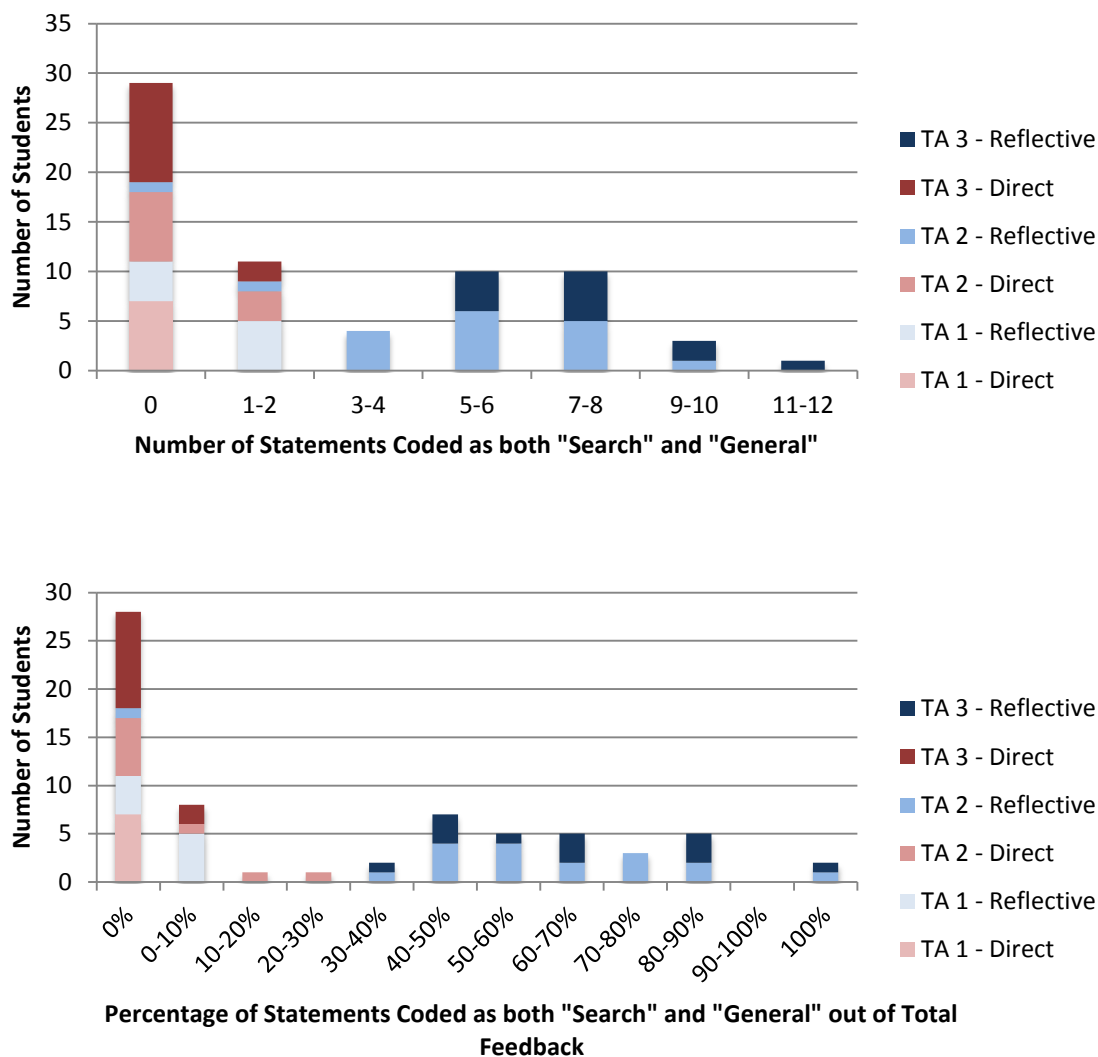


Figure 5.6. Histograms of (a) the total number of reflective statements (coded as both “search” and “general”) received by each student during the semester and (b) the fraction of total number of feedback statements received by each student that were coded as both “search” and “general” – for the “use of evidence to support molecular-level changes” aspect.

These histograms can be used to identify parameters that better define reflective feedback and direct feedback groups that have more clear and pronounced contrasts, identifying students who received primarily reflective feedback on a particular model

aspect and a comparison group that received a similar quantity of non-reflective feedback on that same model aspect.

Based on the above data, narrower reflective feedback subgroups were identified for each aspect, consisting of students who received – over the course of the semester, prior to submission of the Thermochemistry refined model assignment – at least 5 statements coded as both “search” and “general” (therefore averaging at least one such statement every other assignment), and for whom these statements amounted to over 50% of the feedback they received on that model aspect. Therefore, this reflective feedback subgroup consists of those students for whom the majority of the feedback was reflective, and who received the most individual instances of reflective feedback. Percentage of feedback coded as “reflective” was included as a criterion in defining these subgroups in order to ensure that the two groups did in fact see contrasting types of feedback, and not merely contrasting quantities of feedback. For the “discussion of changes to molecular-level model” aspect, 17 students fell into this reflective feedback subgroup; these students received between 6 and 16 feedback statements on that aspect in total, for an average of 9.88 total feedback statements. These reflective feedback statements amounted to 76% of the total quantity of feedback received by these students. Statistics for the “use of evidence to support molecular-level changes” aspect reflective feedback subgroup ($n=19$) are reported in Table 5.6. Note that, although there is some overlap between the two subgroups, the set of 17 students who were classified into the reflective feedback subgroup for the “discussion of changes to molecular model” aspect is not identical to the set of 19 students classified into the reflective feedback sound group for the “use of evidence to support molecular-level changes” aspect.

To more directly compare the effects of receiving reflective feedback to receiving more direct feedback, a comparison group of students also had to be identified. This comparison group for each aspect was drawn from students who received less than 10% reflective feedback on that particular model aspect, but who received a total number of feedback statements within the same range as the reflective feedback group. Therefore, this direct feedback subgroup consists of those students for whom the overwhelming majority of feedback was not reflective, but who also received a substantial number of instances of direct feedback. The range and average number of total feedback statements delivered on each model aspects are reported in Tables 5.5 and 5.6.

However, despite the identification of groups with more pronounced feedback, no clear impact of feedback on performance (as measured by expert rating, as measured by student self-ratings, or in terms of calibration accuracy) can be discerned. No significant differences exist between subgroups with respect to the use of evidence aspect; however, for the changes to molecular-level model aspect, students in the reflective feedback subgroup do rate themselves more highly than those in the direct feedback subgroup ($p=0.003$). In the final model of the semester, although the reflective feedback subgroup rates themselves more highly on average, they also performed better and were better calibrated on the first model of the semester, so their later self-ratings may simply echo their earlier performance.

More direct comparisons could conceivably be made by creating groups matched with respect to early semester measures such as performance on the first model assignment or Toledo placement exam scores, but this would result in unacceptably small groups to compare. A different approach, based on modeling the performance of all students using the available variables, would be more likely to yield meaningful

information, and the possibility of constructing such a model is described in the next section.

Table 5.5. Average expert ratings, self-ratings, and calibration scores for subgroups identified as receiving the most reflective and the most direct feedback on description of changes to their molecular-level models. Reported p-values reflect independent samples t-tests between direct and reflective feedback groups.

“Description of Changes to Molecular-Level Model”	Reflective Feedback Group	Direct Feedback Group	
Group Size	17	11	
Total Number of Feedback Statements	6-16 (avg: 9.9)	8-19 (avg: 12.1)	
Expert Rating: Spectroscopy	0.500	0.273	<i>p</i> =0.079
Expert Rating: Thermochemistry	0.468	0.395	<i>p</i> =0.526
Self-Rating: Spectroscopy	3.824	3.909	<i>p</i> =0.817
Self-Rating: Thermochemistry	4.529	3.636	<i>p</i> =0.003
Calibration: Spectroscopy	-0.034	0.182	<i>p</i> =0.059
Calibration: Thermochemistry	0.068	0.046	<i>p</i> =0.874

Table 5.6. Average expert ratings, self-ratings, and calibration scores for subgroups identified as receiving the most reflective and the most direct feedback on use of evidence to support changes to their molecular-level models. Reported p-values reflect independent samples t-tests between direct and reflective feedback groups.

“Use of Evidence to Support Changes to Molecular-Level Model”	Reflective Feedback Group	Direct Feedback Group	
Group Size	19	17	
Total Number of Feedback Statements	7-15 (avg: 10.31)	7-15 (avg: 11.9)	
Expert Rating: Spectroscopy	0.289	0.147	<i>p</i> =0.195
Expert Rating: Thermochemistry	0.289	0.234	<i>p</i> =0.532
Self-Rating: Spectroscopy	3.579	3.471	<i>p</i> =0.803
Self-Rating: Thermochemistry	4.316	4.353	<i>p</i> =0.879
Calibration: Spectroscopy	0.145	-0.015	<i>p</i> =0.321
Calibration: Thermochemistry	0.219	0.128	<i>p</i> =0.488

5.5 Potential Correlates with End-of-Semester Measures

The same three outcomes as previously discussed – experts’ rating of students performance, students’ self-ratings, and students’ calibration with respect to the instructor’s grade, were selected as factors that may be influenced by the feedback students received. As each of these measures can be studied for each of the five model aspects evaluated, there are 15 possible outcome variables related to the final model of the semester that could be affected by the feedback manipulation. I have focused on the final model of the semester here because of the relatively small number of feedback instances delivered pertaining to each model aspect over the course of the term, hypothesizing that, if an effect of feedback style is present, it is more likely to be seen at the end of the term, after students have been exposed to a greater quantity of feedback of a particular type. Because of the large number of variables being studied, caution should therefore be used when interpreting the statistical significance of any individual correlation.

To develop models for students’ performance on these aspects of their models as a function of feedback received, possible covariates measured within students’ performance prior to the delivery of feedback were examined for possible inclusion in the models. These potential covariates included students’ pretest scores (both the math and chemistry-general subsections of the American Chemical Society’s Examinations Institute 1998 Toledo Chemistry Placement Examination) and the analogous measures (expert rating, student self-rating, and student calibration) for the first model assignment of the semester. There were correlations between these pre-measures and students’ performance in the final model of the semester for 6 of the 15 variables. These

are listed below, and are candidates for inclusion in models describing students' end of semester performance on the relevant measure:

- Expert rating for Thermochemistry molecular-level model:
 - correlated with Toledo exam math subset, $\rho=0.281$, $p=0.021$
- Self-ratings for Thermochemistry description of changes to macroscopic model:
 - correlated with Toledo exam math subsection, $\rho=0.355$, $p=0.003$,
 - correlated with Toledo exam chemistry subsection, $\rho=0.258$, $p=0.003$,
 - correlated with Spectroscopy model expert ratings for the same aspect, $\rho=-0.279$, $p=0.021$.
- Calibration for Thermochemistry description of changes to macroscopic model:
 - correlated with Toledo exam math subsection, $\rho=0.303$, $p=0.012$.
- Self-ratings for Thermochemistry use of evidence to support molecular-level changes:
 - correlated with Toledo exam chemistry subsection, $\rho=0.251$, $p=0.039$.
- Calibration for Thermochemistry use of evidence to support molecular-level changes:
 - correlated with Spectroscopy model calibration for the same aspect, $\rho=0.257$, $p=0.035$.

There are no discernible patterns in the aspects of students' end-of-semester performance that are correlated with various pre-measures, except that students' self-

ratings for two different model aspects (changes to macroscopic model and use of evidence) are positively correlated with the Toledo exam chemistry subsection.

Aspects of feedback may have also affected students' performance; available measures of feedback include:

- the total number of feedback statements a student received on a particular model aspect (this may be somewhat entangled with students' grades; see the discussion below),
- the number of reflective feedback statements (those coded as both "search" and "general") that a student received on a particular model aspect, and
- the fraction of feedback statements that are reflective out of the total amount of feedback delivered.

When broken down by model aspect, a number of statistically significant relationships emerge. Four of these are positive correlations between students' calibration score on a model aspect in the Thermochemistry module and the number of feedback statements received concerning that aspect over course of the semester; this relationship is found for every model aspect except for use of evidence:

- macroscopic model: $\rho=0.273$, $p=0.024$,
- molecular level model: $\rho=0.392$, $p<0.001$,
- description of changes to macroscopic model: $\rho=0.361$, $p=0.001$,
- description of changes to molecular-level model: $\rho=0.340$, $p=0.002$, and
- use of evidence to support molecular-level changes: $\rho=0.126$, $p=0.153$.

Thus, students who received more feedback were had higher calibration scores, reflecting the fact that they were not as well attuned to their instructor's grades on that

particular model aspect. However, this relationship may not be due to feedback alone. As described previously, TA 1, who delivered the most feedback, also assigned the lowest grades, and had more poorly calibrated students, whereas TA 2, who delivered the least feedback, had better calibrated students because there was less of a difference between the instructors' high ratings and students' high self-ratings. Because the grades delivered are likely to be more salient for students than written feedback, it seems unlikely that written feedback is the primary factor responsible for differences in calibration between students who receive varying amounts of feedback. This interpretation is corroborated by two of the other relationships between performance and feedback, which are significant negative correlations between students' self-ratings for two model aspects (macroscopic model and description of changes to molecular-level model) and the number of feedback statements received. As described in Section 4.5, students who received lower grades (and who therefore would likely have also received more feedback from their instructors) gave themselves lower-self assessments, but were also more poorly calibrated.

Only three further significant correlations exist between end-of-semester measures (on student performance or calibration within the Thermochemistry model) and attributes of feedback received over the semester:

- Expert rating for molecular-level model:
 - correlated with number of feedback statements delivered on the same aspect during the semester, $\rho=0.255$, $p=0.036$
- Calibration (with respect to TA rating) for use of evidence to support molecular-level changes:

- correlated with number of reflective feedback statements delivered on the same aspect during the semester, $\rho=-0.309$, $p=0.010$
- correlated with fraction of reflective feedback delivered on the same aspect during the semester, $\rho=-0.260$, $p=0.032$.

Because these correlations do not reveal any patterns common across multiple model aspects, there is not compelling evidence of reflective feedback having significant effects on students' performance, self-assessment, or calibration. In the absence of evidence that feedback type is correlated with end-of-semester measures, it is not logical to attempt a "fishing expedition" to develop a multivariate model incorporating feedback type that accounts for students' performance or self-assessments.

Although the lack of impact on student performance or self-assessment may seem surprising given the pronounced contrasts in the feedback students received, there are several factors that may have contributed to failure to observe any impact of feedback in these analyses.

First, despite the substantial number of feedback statements transcribed and analyzed for this study, there were still relatively few feedback statements actually delivered per student per model aspect. Even for the aspect of students' models on which the most feedback was delivered – presentation of molecular-level models – the average student received 20 feedback statements over the course of the semester on a total of nine refined and four initial models. On use of evidence, despite its importance in scientific explanations, the average student received only 12 feedback comments, barely more than one per refined model.

Second, students may not have attended to the feedback delivered. Anecdotally, one TA in this study recalled an informal experiment in which the instructor wrote

feedback on laboratory reports offering students bonus points if they emailed him. Less than a third of the students took advantage of this offer. Despite statements in which students say they value feedback, many students may not closely read the feedback that is delivered, let alone attempt to recall it in future model-writing attempts – which may come weeks after the previous feedback was received, and which often pertained to different chemical systems. Third, there may have been other forms of feedback that students believed to be more meaningful than their instructor’s written comments. Among these are the grading rubrics used on each assignment, which list each model aspect and identify whether the instructor has graded that aspect “present and complete,” “present but incomplete,” or “not present.” Identifying the parts of the model that are deficient may be as much feedback as students care to process when reviewing their graded work, particularly when this information is easy to come by on the grading rubric. Although all three instructors who participated in this study attached the grading rubric with numerical grades to the back of students’ work rather than the front, anecdotal and video evidence show that students frequently flip directly to the rubric page upon receiving graded models.

However, within this same study population, there exists another possible window with which to examine the effect of feedback: student resubmissions of refined models. When students are motivated to improve a particular piece of writing, the feedback delivered on that assignment is much more likely to be viewed as relevant than feedback from a previous model, and responses to differing types of feedback may be more pronounced and easier to discern.

5.6 Analysis of Resubmissions

During the study described above, students wrote (and researchers analyzed) the refined models written immediately following laboratory experiments. However, for one model within each of the first four laboratory modules completed during the course of the semester, students were also provided with the opportunity to resubmit their work for the opportunity to earn a higher grade. This opportunity was offered for the second refined model in each laboratory module. For each module except for Dissolution, this model was the final model submitted for that module. In the case of Dissolution, although students submitted a refined model after the third week of activities, this was also preceded by an “elaborate your model” assignment that asked students to consider a new chemical situation – a precipitation reaction, rather than the dissolving of two compounds as in the second refined model.

Students were informed in advance, on the course syllabus, that resubmission opportunities would be offered. Each resubmission was due two weeks after the original submission, affording a week for the instructor to grade and return the original assignment with feedback, and a week for a student to make revisions and resubmit the model. Students were only allowed to submit if they had turned in their original work on time. Resubmissions were graded by the instructor, and each student’s final grade on the assignment was the average of the grades earned on the original submission and the resubmission.

For this analysis, because effects of feedback across different model assignments were not examined, it was not necessary to limit the set of students to those who completed every assignment. Consequently, over 80 consenting students were eligible to resubmit each model – but, as shown in Table 5.7, only a fraction of these students did

so, and the percentage of students who chose to resubmit their models fell from over 50% for the Spectroscopy, the first model of the term, to less than 10% for Nanoparticles, the last model on which resubmissions were permitted. Interestingly, for each model after Spectroscopy, a greater fraction of female students resubmitted work as compared to male students; for the Smells model, this difference approaches statistical significance ($p=0.075$, Fisher's Exact Test).

Table 5.7. Numbers and percentages of students who opted to resubmit a refined model for each laboratory module; also broken down by gender.

	Spectroscopy	Smells	Dissolution	Nanoparticles
Students Resubmitting	51% (44/86)	15% (13/85)	19% (17/88)	9% (8/85)
Males Resubmitting	53% (21/40)	8% (3/40)	12% (5/41)	5% (2/40)
Females Resubmitting	50% (23/46)	22% (10/45)	26% (12/47)	13% (6/45)

The analyses of feedback discussed above examine the potential effects of feedback from one model to the next. In other words, does feedback on an aspect of one model have an impact on the quality of students' work in a later model, on a parallel aspect? Applying feedback from one model to the next requires a student to transfer the knowledge gained into a new context. Even though the instructor may see the parallels, they may not be as clear for the student.

In looking at the effects of feedback from a student's original submission to a resubmission, less abstraction is needed in identifying parallel tasks to compare students' performance. Indeed, as much of the text of a resubmitted model is often the

same as the text of the original model, it is straightforward to identify both the differences between the two texts and the feedback statements that relate to the revision.

Therefore, analysis of resubmissions proceeded in a different manner than the coding of models described above. First, a student's original model and resubmission were read in parallel to identify any differences between the two. Any statement that was revised, added, or eliminated was noted.

As described within the expert coding scheme, several aspects of students' models build upon other aspects. For instance, if a student describes a behavior of molecular-level particles in their refined models, then they should also describe how their model of that behavior has changed since their initial model. Therefore, if a student adds a new facet to their description of molecular-level behavior in their resubmission, they should also add discussion about whether and how that aspect has changed since their initial model. Similarly, new descriptions of changes to a molecular-level model should be supported by citation of evidence.

For example, student 08-045, in her original submission for the Spectroscopy module refined model, writes for her description of molecular-level changes:

“I changed quite a bit of my initial thoughts for my refined model. In my initial model, I made predictions about heat and energy rather than what was being absorbed and transmitted. For my refined model, I left out theories about heat and energy and focused more on molecules and how they absorb or don't absorb certain wavelengths of light. I guess one example of experimental evidence I used to change my initial model is that during the experiment I had no way to measure energy (or heat energy). Also, I was able to refine my

predictions and hypothesizes [sic] by actually collecting data, such as the % of transmitted light, using the machine.”

Her TA (TA 2) awarded her full credit for a description of “key changes to molecular-level model of what happens when you shine light on the sol’ns,” (emphasis added); however, no specific evidence is offered in support of her ideas related to these new ideas. With respect to the lack of evidence, the TA offers the reflective feedback: “Have you consistently identified data that supports your refined model? Have you backed up all of these changes w/ evidence?”

The other major aspect of the model that the rubric asks the student to address is “key changes to molecular-level model of dilute and concentrated FD&C blue #1,” (emphasis added); this is absent from the model, and no credit is given for this by their TA. In her refined model, this student begins by essentially restating the initial sentences from her first submission:

“I changed quite a bit of my initial model to make my refined model. In my initial, I made predictions about heat and energy rather than what was being absorbed and transmitted. For my refined model, I focused on molecules and how they absorb or don’t absorb certain λ of light.”

She then goes on to provide new content, describing a new idea relating to the contrast between the concentrated and dilute solutions, as well as citing evidence to support this change to her model:

“Blue Dye #1 concentrated always absorbed more light because it has more molecules in it than Blue Dye #1 diluted. In fact, the concentrated absorbed 0.45 while diluted absorbed 0.23 at 400 nm (for example). This is an example how the a [sic] solution with more molecules will obviously absorb more photons (one change).”

Note that not only has this student identified a change to her molecular-level model that was not described in her original model, but also identified evidence to support the change. The TA, in grading the resubmission, awards full credit for description of changes to this part of the model, as well as for supporting evidence.

The student goes on to elaborate on how her model has changed with respect to the interaction of light and dye molecules:

“Also, I changed my model on what happened when light is shined on a solution. I decided that when a light is shone on a solution, the color (λ basically) of the light determines how much the solution will absorb.”

This is the original aspect of molecular-level changes that was not supported by evidence, prompting the TA to write the original feedback statement. The student responds to the feedback by adding evidence to support the previously unsupported aspect:

“For example, the blue dyes are blue, so they absorb blue light the least of all. (In the experiment the absorption was as low as .0021 for blue dilutes and .0031 for concentrated.)”

In revising her model, this student has not only offered evidence to support an aspect of her model that was previously unsupported and whose absence prompted instructor feedback, but has offered evidence to support an entirely new segment of her model. One can easily imagine how this might be prompted by a TA feedback comment: “Have you backed up all of these changes w/ evidence?” as the student decides what evidence to offer in her refined model.

It might be argued that the student also had the categories on the grading rubric to point out this issue, but students often fail to support aspects of their models that are

added. For example, a student who adds a molecular-level aspect to their model should then discuss how this aspect of their model has changed (or not). Student 08-005's model illustrates an example of how this can happen.

In his first submission, he writes, for his molecular-level model and changes section:

"Molecular. Before the light is shone on the dye molecules, the molecules will be moving freely around the tube. When light is shone on the molecules some of the light will be absorbed. The amount of absorption will depend on the wavelength of the light. When the wavelength corresponds to blue light almost all of the light will be absorbed by Blue #1 concentrate and Blue #1 dilute molecules. When the light is absorbed the electron in the dye molecules will be excited and move to a higher energy level."

"Changes. In both my initial macro and molecular models I expected light to be emitted by the molecules. However, this was not true, either the molecules absorbed light or let it pass through, but never did the molecules emit light. The only part of my initial model that stayed the same was the pre-light section."

This student offers some description of a molecular-level model of what happens when you shine light on the solutions (receiving half-credit from TA 1 for this aspect), but has no description of the molecular-level contrast between dilute and concentrated dye solutions. In response to this, the TA offers the following feedback on the molecular-level portion: "Diff btw conc & dilute?", as well as feedback about the partially present discussion of the effects of wavelength: "What happens with other wavelengths? Why is there a difference for different λ s from a molecular-level perspective?"

With respect to changes, the instructor awarded half credit for description of changes relating to "what happens when you shine light on the soln's," but no credit for

description relating to the students' "model of dilute and concentrated FD&C Blue #1." The instructor offers the following feedback on changes: "What were the key aspects that remained the same? Identify them. It appears to me that some of your 'after light' ideas remained the same?" This TA's written feedback is specifically targeted toward remedying the deficiencies in the discussion of changes rated as "present but incomplete," (those pertaining to interaction of light and solutions); the aspect that is omitted entirely is identified on the grading rubric as "not present," but not specifically mentioned within the instructor's written feedback.

The student's resubmission expands on both aspects of his molecular-level model – both the contrast between dilute and concentrated solutions and the interaction of light with the solutions:

"Before the light is shone on the dye molecules in the solution, the molecules will be moving freely around in the test tube. When light is shone on the solutions some of the light will be absorbed by the molecules. The amount absorbed will be different between the concentrate and dilute Blue #1. The concentrated dye molecules will be more closely packed together and there will be more of there. This will mean more light will be absorbed. The dilute on the other hand will have few molecules packed less closely and will absorb less light. When the light is at a wavelength that is anything but blue more light will be absorbed by the blue dye molecules. When this light is absorbed the electrons in the molecules will be excited and move to higher energy levels."

The student clearly expands upon his discussion of both the differences between the concentrated and dilute solutions, as well as about the effects of wavelength on the absorption of light. The student's grades improve to 4 out of 4 points for presenting both

aspects of the model. This is a dramatic improvement on his discussion of concentrated and dilute solutions, which were not specifically described in the original submission. Both of these discussions, though, offer substantial ideas that are new to the resubmission and which should then be discussed in the student's "changes" section.

In his resubmitted discussion of changes, the student writes:

"In my initial model I thought that the molecules would emit light. However I learned that they don't. In reality the dye molecules all ways absorbe [sic] some of the light shone on them. Some things that did not change were my description of what the test tubes would look like before light before light was shone on them, and the fact that the electrons in the molecules will be excited by the energy from the light. My thoughts on how light would interact with the dye molecules changed because of the experiment that we did. I thought that the molecules would emit light, but in the experiment we found that when light was shone through the solutions the light that came out the other side was weaker than the light that was shone in. The concentrated solution was harder for light to pass through than the dilute solution. I think this is because there are more molecules in the concentrate and these molecules absorbe [sic] the light causing less to pass through."

This model segment includes substantial expansion of discussion of changes. In particular, the student identifies ways in which his model about dye molecules interacting with light does not change. In doing so, the student responded to the original, direct feedback regarding describing ways in which their model had stayed the same. However, they do not address a major aspect of their molecular-level model regarding the interaction of light with dyes – the fact that different wavelengths of light interact with the dye differently. The instructor did not provide feedback in the original submission regarding discussing changes to this facet of their model, presumably

because it was not part of the original molecular-level model. The student's description of changes regarding the interaction of light and dye, although it responds to the feedback that was given, is still not complete, and the student receives only half credit from their TA.

Furthermore, the student should have also discussed changes to a new aspect of their molecular-level ideas that was incorporated into their resubmission – the description of dilute and concentrated dye solutions. However, although he mentions the concentrated and dilute solution within the resubmitted “changes” section, he does not discuss how this aspect of his model has changed. The TA therefore only awards 1 out of 6 possible points for identifying changes to the molecular-level model of dilute and concentrated solutions (presumably giving minimal credit because, although this aspect is mentioned under the “changes” heading, changes are not clearly discussed or identified). By failing to discuss how this aspect has changed, the student has failed to support an aspect of his model that was added in his resubmission.

Student resubmissions within the Spectroscopy module were analyzed in a manner similar to that described above. First, models were compared side-by-side to identify differences between students' original submissions and resubmissions. These differences were summarized and identified by the aspect of the model in which they appeared. Any feedback statements pertaining to altered portions of the models were also transcribed at this time.

Following this, differences were examined to evaluate whether they constituted new ideas that should be supported in another section. In particular, these included:

- new aspects of a student's macroscopic model that should be discussed in identifying how their macroscopic model has changed,

- new aspects of a student’s molecular-level model that should be discussed in identifying how their molecular-level model has changed, and
- new identifications of ways in which a student’s molecular-level model has changed (or stayed the same) that should be supported by citation of specific experimental evidence.

If any new aspects of a student model were included that should be supported, I then identified whether another addition in the revised model provided that support. I also identified whether there was written feedback present that pertained to this supporting aspect, and, if so, what type – reflective or direct. For example, in the first example given above, the student saw reflective feedback relevant to providing supporting evidence, whereas, in the second example, the student saw direct feedback related to completely describing changes to a molecular-level model.

In this case, reflective feedback is defined as that which asks students to search an aspect of their model to determine whether it is complete – the same criteria which defined the “search” characteristic of the feedback coding scheme described previously). Direct feedback addresses a deficiency or a failure to provide support, and may ask a student to provide such information, but does not ask a student to evaluate whether or not that information is present.

	Direct Feedback	Reflective Feedback
Changes to Macroscopic Model	<ul style="list-style-type: none"> • “What were the key aspects that remained the same? Identify them.” • “What do you think about this now?” • “What aspects of your model stayed the same, or did it all change?” • “Describe your changes to your models of concentrated and dilute sol’ns.” 	<ul style="list-style-type: none"> • “Have you described changes to your model dilute/conc and reacting with light?” • “Have you identified whether each aspect of your model is new or changed?” • “Is each identification of changes specific? Have you discussed changes for each major aspect of your model?”
Changes to Molecular-Level Model	<ul style="list-style-type: none"> • “What, specifically, did you add?” • “Molec level changes conc/dilute blue & expt evidence conc/dilute & how it reacts w/light” • “Any aspects same?” • “Key change to molec dilute vs. conc” • “Key aspects that remained the same, w/evidence?” 	<ul style="list-style-type: none"> • “Are all your changes specific?” • “Have you described changes for both macro & molec aspects of dilute/conc blue dye?” • “Have you identified changes for each part asked or in the assignment?”
Use of Supporting Evidence	<ul style="list-style-type: none"> • “Identify which evidence, specifically, supports this, and explain how.” • “What evidence do you have to support this?” • “Specific exp evidence?” • “Specific experimental evidence to support molecular-level model?” 	<ul style="list-style-type: none"> • “Do you consistently identify data that supports your refined model?” • “Have you backed up all of these changes with evidence?” • “Have you consistently identified data that supports your refined model? Have you backed up all of these changes w/evidence?”

Figure 5.7. Examples of direct and reflective feedback statements within Spectroscopy model resubmissions analyzed.

In all, within the 45 resubmissions analyzed, there were 77 statements identified that were new to the resubmissions and should be supported within another aspect of the model – 21 new macroscopic model ideas, 29 new molecular-level model ideas, and 27 new molecular-level model changes. Table 5.8 presents the fraction of model aspects that were supported, broken down by model aspect and by type of feedback left. (Note that the counts here refer to specific types of model revision, and that, because multiple

revisions could occur within students' models, some students' models contributed multiple instances to this data set, while others contributed none. Individual student characteristics are not taken into account in this analysis.)

Table 5.8. Fractions and numbers of instances in which an idea added to a student's model in the aspect indicated is supported in the model aspect indicated in parentheses, broken down by the type of feedback present.

	Overall	Macroscopic/ Changes to Macroscopic	Molecular/ Changes to Molecular	Changes to Molecular/ Use of Evidence
Reflective	73% (19/26)	86% (6/7)	63% (5/8)	73% (8/11)
Direct	37% (14/38)	50% (5/10)	20% (3/15)	46% (6/13)
No Feedback	15% (2/13)	25% (1/4)	0% (0/6)	3% (1/3)

For every aspect, when reflective feedback was left on the supporting section, the student supported the new aspect appropriately between 63% and 86% of the time. When direct feedback was left on the supporting section – prompting the student to improve or add support for an already-existing aspect, but not asking the student to consider the level of support generally, students provided support between 20% and 50% of the time, depending on the model aspect in question. Finally, when no relevant feedback was left on the supporting section, then students included support in that section for new aspects 25% of the time or less.

Collectively, in a significantly higher fraction of instances, when reflective feedback was provided in a portion of a model, new aspects were supported, as compared to when direct feedback was provided. (Fisher's exact test, $p=0.003$) or than when no feedback was provided ($p=0.002$). The contrast between instances when direct

feedback was given and those where no feedback was given is not statistically-significant ($p=0.304$).

Because of the smaller populations involved, Fisher's exact test has relatively poor power in determining whether there are statistically-significant differences between the categories for individual model aspects. None of the contrasts between reflective feedback and direct feedback subsets or between direct feedback and no feedback subsets show statistically significant differences when broken down by particular model aspect.

5.7 Discussion

Analysis of the impacts of "direct" versus "reflective" feedback reveal that there is little discernible impact on students' performance or calibration from model to model across the semester. However, analysis of student resubmissions suggests that reflective feedback may have an impact in a situation in which students are more likely to be attuned to feedback.

The contrasts between direct and reflective feedback may simply be too subtle to have a substantial impact on student performance between successive model assignments. The distinction between "Include evidence to support this molecular-level change." (a "direct" statement that does not ask students to examine their model) and "Have you included evidence to support each molecular-level change?" (a "reflective" statement which does ask students to examine their model) may not be relevant to a student who simply wants to know what portions of an assignment they were marked down for. Other characteristics of written feedback not captured in these analyses may have much larger impacts on student performance. However, as a reminder to check

every aspect of their model thoroughly in the context of revising and improving it, “reflective” feedback statements may be more effective than “direct” feedback statements at prompting more accurate evaluation of one’s writing and therefore higher quality resubmission.

The effects of particular written feedback comments across multiple assignments, if they exist, are likely to be smaller in size and may require larger studies to be measurable. Furthermore, effects of feedback on self-evaluation (in this study, measured in the form of calibration scores) may be more pronounced in contexts such as resubmissions. Given the significant relationships between feedback type and changes made in resubmissions, it would be reasonable to expect differences in monitoring accuracy for students resubmitting their models after having received different types of feedback. Student self-assessments for performance on resubmissions were not captured in this study, but examining such metacognitive judgments in this context would be a logical continuation of this work.

Finally, just as the introductory modeling and critiquing exercises described in Chapter 3 may have improved students’ later model-writing by developing more accurate self-assessment, certain types of feedback may similarly have impacts on the quality of later models if students attend to them closely during the process of revising a model for resubmission. Indeed, when instructors provide feedback to students, or when researchers attempt to study the effects of feedback, it is prudent to provide an incentive for students to attend to written feedback closely – otherwise the effects of that feedback may be difficult to discern.

Chapter 6: Analysis of Student Reflection from Video

The previous chapters have, in part, examined the impact of various aspects of instruction on students' written models, with particular attention on the quantitative assessment of their performance on several key aspects. In this chapter, I choose to take a more qualitative approach, centering on analysis of video of students working in the laboratory, to examine how students arrive at some of the ideas exhibited in their refined models. Here I attempt to go beyond simply quantifying student performance as an outcome measure. I am interested in not just whether and how students come to develop models of a particular quality, but in how students come to express particular ideas that appear in their refined models.

It seems reasonable to assume that students' refined models reflect, at least in part, the activities, reasoning, and discussions that they participate in or witness during the laboratory period. This window into students' thinking is necessarily incomplete – even examining the discussions that students engage in during the laboratory period does not grant a researcher direct access to the internal cognitive processes that may in fact be the crucial link in developing a deep understanding of a topic – as Chi notes, “overt activities cannot accurately discriminate among... cognitive processes definitively.”⁷⁹ However, in-class discussions can offer an additional perspective on students' thinking processes. In these discussions, we can see how students respond to a range of situations, pieces of experimental evidence, or questions. However, like written work, discussions also have limitations as a source of evidence of cognition in that students often describe things in briefer or vaguer terms than they do in writing.

Past research on the MORE Thinking Frame has shown that, when using curricula or instructional approaches that are effective at generating a deeper understanding, students engage in deeper discussions during the class or laboratory period. Rickey^{10b} found that instances of high-quality reflection, even in implementations of the MORE Thinking Frame at UC-Berkeley, were relatively rare, but were much more common than in a control classroom using more traditional instructional methods.

In this chapter, I investigate both the factors that instigate episodes of reflection and their effects on students' later written models. My primary questions here are: *What activities or prompts within the laboratory period foster reflection related to key aspects of students' models? What are the relationships, if any, between during-class reflection and the content or quality of students' post-laboratory written refined models?*

6.1 Selection of Video Groups and Data Collection

To closely examine the ideas expressed by students throughout the laboratory period, I chose to capture in-class video. Video offers opportunities unavailable in naturalistic observation to closely parse and repeatedly review the discussions that occurred and the contexts of those discussions.

Initially, five groups of students were selected for video recording and analysis, one group from each laboratory section selected for study. Video groups were constructed by the researcher prior to the third laboratory period of the semester, and students were directed to work in those groups throughout the term. Students in this laboratory course routinely worked in groups of four, so groups of that size were

constructed using stratified random sampling from the class, selecting one student from each quartile (top 25%, second 25%, etc.) as measured by scores (sum of both subsets) on the Toledo pretest. One additional criterion was applied: that no group could consist of three male students and one female student, building off of research indicating that groups containing a minority of female members tended to marginalize the contributions of that member.⁸⁰

Following the third week of the semester, TA 1 noted that the students in the video group selected within that instructor's class appeared less engaged in class activities than other groups. This instructor noted, for example, the failure of three of the four to complete both pages of a required safety worksheet. Therefore, the instructor and researcher decided to record an additional video group within that laboratory section. As groups had already formed and begun working together during the first weeks of the semester, a group that had already self-assembled was chosen. This group was therefore not balanced by performance, and consisted disproportionately of high-performing students.

Video of each laboratory group was recorded every week during the semester. Cameras were focused on the laboratory group and a wireless microphone was placed in the middle of the laboratory group's table. Students were asked explicitly for consent to be video-recorded during the first week of the laboratory course, and were compensated \$10 for each laboratory period during which they were recorded, with payment made after the conclusion of the course.

For this study, I have focused on two laboratory activities. The first of these is the first week of the "What Happens When Chemical Compounds Are Added to Water?" module. As outlined in Chapter 2, in this experiment, students use conductivity as a

probe of the molecular-level behavior of different substances that are added to water, including mixtures of water with ionic substances (NaCl, CaCl₂, etc.), covalently bonded substances (C₁₂H₂₂O₁₁, C₆H₁₂O₆, etc.), and metallic substances (Zn). For each, students record observations and measure the conductivities of the resulting mixtures. In this implementation of the laboratory experiment, semi-quantitative conductivity testers were used. On these testers, the flash rate of an LED indicates conductivity (on a roughly logarithmic scale). Students were instructed to count the number of flashes that occurred within a period of 30 seconds, and therefore often reported measurements in units of “flashes”. Students then predict and measure the relative conductivities of solutions of 0.01 g of NaCl and 0.01 g of NaI dissolved into 100 mL of water, and finally, after finding that the NaCl solution has higher conductivity, attempt to predict what mass of NaI they would need to dissolve in 100 mL of water to create a solution with comparable conductivity.

This laboratory module was chosen in part because it has been demonstrated in previous literature to result in a significant number of students exhibiting more scientifically correct ideas in their refined models as compared to their initial models,¹² and because it has been held up as a “model” module with which students can be introduced to the MORE Thinking Frame. Preliminary analysis of student video also revealed that student discussion in this laboratory module contained a greater number of episodes of student reflection about molecular-level ideas than other laboratory modules examined, such as the Thermochemistry module, which is the subject of the second case study, and is described in Section 6.3.2.

In both of these experiments, consistent with the approach taken by the MORE curriculum, students are not directly told what is happening on the molecular level, but

are asked in both the post-laboratory refined model assignment and in reflection questions embedded throughout the experiments to think about molecular-level behavior and its connections with experimental evidence.

6.2 Factors Instigating Molecular-Level Reflection

Based on the methodology of identifying “action-relevant episodes” described by Barab et. al,⁸¹ analysis of video began by dividing the recording into segments, each of which were delimited by a change in participants, activity, or topic of discussion. In the first pass through the video, episodes related to students’ discussion of molecular-level ideas, discussion of student models, and students’ attempts to make sense of evidence, were tagged as candidates for further examination. Episodes were also tagged based on the participants involved in the interaction – group members alone, group members conversing with the instructor, or whole-class discussion– to facilitate this later analysis. Initial review of these data suggested that reflection about the molecular level was far more common across laboratory groups than other types of reflection directly relevant to students’ models. Therefore, in a second pass through the tagged video, segments specifically relevant to discussion of the molecular-level were identified, using the same criteria for defining molecular-level as was used in the analysis of student models described in Chapter 3.

In total, 47 episodes of reflection on the molecular level were coded within the six laboratory videos analyzed. For each episode, the remark or incident that initiated the episode, changing the topic from the previous discussion or initiating discussion after a pause in conversation, was identified. Based in part on the context identified in the first-

pass coding, the sources of the initiating remark could be grouped into the following categories:

- An instructor's question to the lab group – 9/47 instances (19%)
- An instructor's question to the entire class, resulting in an interaction that the group attends to or participates in – 19/47 instances (40%)
- A reflection question embedded in the laboratory manual considered by group members – 8/47 instances (17%)
- Discussion between group members – 11/47 instances (23%; this category includes those episodes that could not be identified as being sparked by a question from the instructor or laboratory manual, but may have arisen due to experimental observation, reflections by group members, etc.)

Most notable here is the observation that the vast majority of instances of molecular-level reflection were prompted directly by the instructor or laboratory manual. Less than one quarter of the episodes of molecular-level reflection originated without a direct prompt from a source external to the group. This highlights the important role played by the instructor and the reflection questions within the laboratory manual in shaping the ideas students think about during the laboratory.

As important as who provides the prompt that causes students to reflect about the molecular-level is the content of that prompt. To uncover patterns in these types of questions, I transcribed and examined this set of statements and grouped them into categories, refining these groupings following discussion with other researchers, who helped identify recurring themes within the data

Categories of instigating questions or prompts are:

- Questions asking directly about the molecular-level (12/47)

- Instructor: “So, what do you think is going on in isopropanol on the molecular level, then?”
 - Lab Manual: “How can you explain this observation in terms of molecular-level behavior?”
- Questions asking for identification of or explanation of patterns across multiple observations (11/47)
 - Lab Manual: “Any trends or patterns that you notice?”
 - Student: “So... the compounds that had nonmetals all had low conductivity.”
- Questions asking for explanation of individual observations (3/47)
 - Instructor: “I was just wondering what you think the conductivity measurement of isopropanol tells you.”
- Statements making and justifying predictions (12/47)
 - Student: “So I’m guessing... this is an ionic compound, so maybe there will not be much conductivity.”
 - Instructor: “Folks who said that sodium chloride would have a higher conductivity – why?”
- Questions asking about chemical terminology (4/47)
 - Student: “Electronegative – it means what?”
- Statements acknowledging unexpected experimental evidence (2/47)
 - Student: “There goes that theory!”
 - Student: “Forty-two... that’s weird.”
- Other, uncategorized statements (3/47)

I also wished to contrast prompts that fostered molecular-level reflection with those that did not. To that end, I examined the entire set of reflection questions from the laboratory manual, as well as the set of statements that prompted interactions between instructors and students, but did not lead to molecular-level reflection.

With respect to reflection questions in the laboratory manual, all laboratory groups had the opportunity to consider the same questions. I therefore tallied which of these reflection questions prompted reflection on the molecular level for the six recorded groups, and which did not. For instance, questions that prompted-molecular level reflection for more than one group included: “How can you explain this observation in terms of molecular-level behavior?” and “What determines the number of LED flashes? [the conductivity reading]” Both of these fall into categories outlined above – asking for explanations of observations. By contrast, questions that did not prompt molecular-level reflection in the 6 video groups during class tended to focus on requests for recapitulation of observations: “What types of changes did you observe as you added different substances to water?” or “How did the different substances added to water behave when you tested them with the conductivity tester?”

Although instructors interacted with different laboratory groups in different ways, it is also similarly possible to compare the statements made by instructors that prompted molecular-level reflection with statements made by instructors that initiated a new conversational episode without molecular-level reflection. Statements that prompted molecular-level reflection included questions about predictions (e.g., “Is that what you expected to see?”) or about the underlying behavior of chemical systems (e.g., “I was just wondering what you think the conductivity measurement tells you,” “How are you explaining some of the trends for the first part?”) Comments that did not

prompt any molecular-level reflection tended to center around description of observations (e.g., “Look at the water closely,” “You’re making good observations of the solids.”) or laboratory procedures (“Can you start weighing out some of this stuff?” “Can you guys put your DI [water] reading up on the board?”)

Although it seems intuitively obvious that prompts that directly ask about the molecular-level are more likely to prompt discussion about molecular-level behavior than those which do not, it is important to note that prompting reflection about molecular-level behavior is not always a key goal of chemistry instructors in the laboratory and particulate descriptions of matter are not always explicitly discussed in laboratory exercises.⁸²

6.3 Case Studies of Student Discussions

Having identified that discussion of molecular-level ideas and behavior occurred with some frequency within student-student and student-instructor dialogue, I chose to adopt a case study approach to describe whether and how these ideas and discussions, as expressed during the laboratory period, were related to ideas expressed in students’ models.

The methodology used to develop these case studies was as follows. First, for each student participating in a group selected for video analysis, their initial and refined models – written before and following that week’s lab period – were examined, and, for each, statements relating to the molecular-level (as identified in descriptions of molecular-level coding from Chapter 3) were identified and summarized.

A similar process was carried out with respect to dialogue captured in laboratory video, both from discussions students from the video groups participated in and from

whole-class discussions to which those groups attended. Episodes of molecular-level reflection were identified and transcribed, with additional notes about the surrounding context and the point during the laboratory period where the episode occurred. A listing of molecular-level ideas, similar in format to that developed from students' models, was produced from this transcription. This then allowed me to compare the molecular-level statements made in initial and refined models with those stated during the laboratory period.

Finally, once each student's statements and models had been analyzed in this way, I used these data to develop a narrative of the ideas each student expressed before (from their initial models), during (from their discussions video), and after (from their refined models) the relevant laboratory session.

Here, I describe two case studies, first by outlining the sets of ideas that students discuss during the laboratory period, and interspersed with that, describing how these ideas relate to students' initial models. I then examine students' refined models and describe how they relate to the ideas shared during the laboratory period. All student names used in these narratives are pseudonyms.

6.3.1 Case Study 1: “The High Lattice Theory”

The first case study that I describe comes from the classroom of TA1 – in particular, from the group of high-performing students described in Section 6.1. The four students in this group engaged in a greater number of discussions about molecular-level behavior than any other group selected for video analysis.

The first episode of molecular-level reflection begins as students are in the midst of testing the conductivity of their first set of substances added to water:

Instr: So what's the next one you're gonna do?

Jack: [to Marie] I guess we now go down the list and do calcium chloride.

Marie: Go ahead and do that first, so.... See calcium chloride... I've heard that chlorine is kind of... it's... uh...

Deb: [interjecting] ... and the stirrer thing has been rinsed, so you can use it.

Instr: Chlorine is kind of what?

Marie: More... electro... conductivity.

Instr: So...

Kurt: Because the calcium is a 2 plus, so... [looks at and stirs mixture] I think it'll be pretty conductive because there'll be three... three ions per molecule in the solution to conduct electricity.

Instr: How does it compare to the isopropanol, do you think?

Marie: More blinks.

Kurt: More blinks.

Instr: More blinks – why?

Deb: Yeah.

Marie: Umm... because water was about five and a half... and since the isopropanol is the same as water, I think that uh... it'll... there'll be more blinks in the calcium chloride than there were in...

Deb: Well especially because the isopropanol, wasn't that, like, a liquid to begin with, or at least, it wasn't as big of... crystals as...

Instr: So what do you think... what do you think it is that causes the conductivity?

Kurt: It's the number of ions in the solution.

Instr: The number of ions in the solution?

Marie: So...

Instr: Does everybody agree with that?

Marie: It kinda makes sense.

Deb: It would make...

Instr: Why does it make sense?

Marie: Because electricity is negative.

Instr: Electricity is negative, so...

Marie: So... if there's more negative anions in a solution, then... there'd be more electricity.

In this episode, Kurt confidently responds to the instructor's question by offering an idea very close to that that the laboratory experiment is designed to develop – that the number of ions controls the level of conductivity (in actuality, it is the concentration of ions).

At this point in the laboratory module, Kurt has articulated the clearest idea about the molecular-level origins of conductivity. His initial model is the only model of those in this group to provide a scientifically correct description of the dissolution of salt; his only inaccuracy is a reference to the breaking of covalent bonds within sugar crystals, but he also describes sugar after dissolution as consisting of “molecules”

“After: the water molecules surround the salt ions, causing them to break apart from each other:

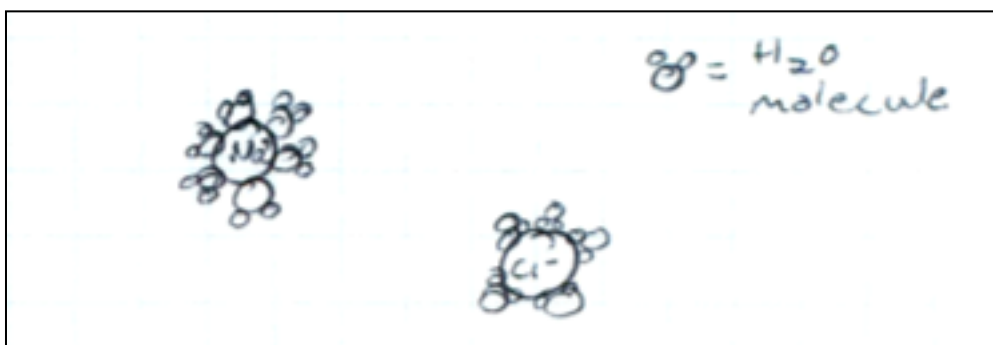


Figure 6.1 Kurt's molecular-level picture of dissolved salt in his initial model.

The water molecules form hydrogen bonds with the sugar molecules, breaking apart the covalent bonds in the sugar crystals.”



Figure 6.2 Kurt's molecular-level picture of dissolved sugar in his initial model.

In contrast with Kurt's confidence, Marie appears to be attempting to develop her own explanation for what causes the conductivity of solutions. Her initial model contains some incorrect ideas about dissolution – particularly that salt, when dissolved in water, will remain as NaCl molecules, but her initial model also contains a relatively sophisticated understanding of the molecular-level causes of dissolution in general:

“When the substance is put in water, the negative parts of water will be attracted to the positive parts of the substance and vice versa. Water molecules will form a shell around these molecules based on charge and will draw them away from the rest of the crystal, resulting in breakdown of the crystalline structure.”

Following the initial episode quoted above, the instructor then probes more explicitly about isopropanol, starting a conversation that involves several other group members.

Instr: But... what... I was just wondering what you think the conductivity measurement tells you.

Marie: I... don't know.

Jack: That... when it mixes with water, it's not really having much of a reaction.

Deb: There's not a lot...

Jack: They're just kind of in there just intermingling, but not really creating any kind of reaction, so they're not creating any energy. So it'd be a different thing than...

Deb: There's not a lot of... increased activity or anything.

Jack: Yeah, you're not going to get much of a different reading.

Deb: The H₂O's aren't working that hard to pull apart the isopropanol.

Instr: Okay.

While Jack and Deb appear to be developing an idea connecting the presence of a reaction or activity within the solution with conductivity, in the following discussion, Marie attempts to try to merge this idea with the ideas about salt being broken apart in water, an idea that she expressed in her initial model (Jack, Deb, and Marie all articulated ideas in their initial models about salt being broken apart into neutral molecules when dissolved) :

Marie: So what was the consensus on the isopropanol?

Jack: That it's not really... really has no has no charge and isn't going to be adding anything to the deionized water. 'Cause of what that little battery thing told us...

Deb: I don't think the H₂O molecules are working as hard to pull it apart. Like...

Jack: It's already... It's not really doing it with the isopropanol. I don't know if that means...

Kurt: What?

Marie: Water doesn't have to work as hard to dissolve... so it'll just kind of mix, and there's really no change.

Kurt: So the water doesn't really dissolve it?

Jack: They just sort of mix together...

Marie: [after a pause] So the more water has to work to break apart something, the more charge there will be?

Jack: Sure

- Kurt: Say that again?
- Marie: So are we thinking that maybe... the more the water has to work to break apart something... the more charge there will be?
- Jack: That sounds... better than anything we've got otherwise, so...
- Kurt: Say that again?
- Marie: So maybe the more... umm... the harder... like maybe the higher the lattice energy, the harder the... the... water molecules have to work to break apart something... the more charge there will be, since... I'm not sure why, but maybe because... I don't know, unless you have another...
- Kurt: So the higher the lattice energy, the... yeah, it's like the more water molecules surround each ion...
- Marie: ... to break it apart, so... more conductivity... more electricity.

In the episodes above, both Deb and Jack articulate the idea that the level of conductivity is related to the processes involved in separating dissolving species. Jack couches this in terms such as “not adding anything to the deionized water,” or “not having much of a reaction,” and, on several occasions affirms such ideas when expressed of group members. Deb is the student who first brings up the structure of the pure compound (liquid as compared to crystals for isopropanol) as a factor; she later refers to “increased activity” associated with dissolving, and describes in greater depth how this might be related to conductivity: “The H₂Os aren't working that hard to pull apart the isopropanol.” Marie offers further elaboration on the idea, introducing the concept of “lattice energy.”

As is intended throughout a MORE laboratory module, further evidence causes members of the group to question the validity of this preliminary model. That evidence comes in the form of their measurements of the conductivity of a dextrose solution; dextrose is a crystalline solid, much like NaCl or many ionic compounds.

Marie: There goes that theory. Darn it.
Deb: So there goes the high lattice theory?
Kurt: It's the... umm... ionic compounds are really conductive.
Marie: OK, so maybe...
Kurt: And the other ones are...
Marie: Maybe because dextrose... Oh, stop. Maybe because dextrose doesn't have any... umm... ionic lattice. It's just a bunch of crystals... or not crystals, but...
Deb: [after a pause] So what were you saying?
Marie: So maybe dextrose is not an ionic compound. Doesn't have to do as much work.

Once again, Kurt offers his model that the salient feature relating to conductivity is that the compounds in questions are ionic. This model is not rejected by the other students. In fact, ultimately the students all revise their models to describe salt breaking apart into ions in solution. However, the students also do not adopt Kurt's explanation for the conductivity levels of the solutions, and instead elaborate and adapt their previous ideas about work done to separate compounds contributing to conductivity.

During a whole-class discussion that follows the previously-presented excerpt, Marie elaborates on her ideas about ionic compounds breaking apart being associated with energy and, therefore, conductivity: "Maybe it means that you have to break apart ionic compounds..." "For dextrose, maybe one molecule of $C_6H_{12}O_6$ gets separated from another one." "...it takes energy to break a bond... then maybe that's why the ionic compounds have that conductivity." As intended, the instructor simply solicits students' ideas, and does not indicate whether any are correct or incorrect.

Following this discussion, students are prompted to predict the relative conductivities of a 100-mL solutions containing 0.10 g of NaCl or 0.10 g of NaI.

(Because of the larger mass of the iodide ion as compared to the chloride ion, the NaI solution contains fewer ions and therefore registers a lower conductivity). The students' discussion of their group prediction proceeded as follows:

- Marie: All right. What do you guys think?
- Deb: I think they'll have the same.
- Marie: Why?
- Kurt: I think... the same, because it's the same number of ions in solution.
- Deb: And, well, and I also said, in the, like, non-metal elements, so, like, the iodine and the chlorine, see...
- Kurt: Yeah.
- Deb: ...they're in the same in the same group.
- Kurt: Mmhmm
- Deb: So it's not like one's...
- Kurt: Yeah,
- Deb: Yeah.
- Kurt: Yeah, it's like the same valence electron configuration and all that.
- Marie: What do you think about the difference of size?
- Jack: That electronegativity... will conduct it a little bit... you know, it could take more energy to pull them apart, because one's more electronegative and therefore you create more of a reaction.
- Marie: Yeah, that's what I think.
- Deb: So you guys think that the size is going to affect...
- Marie: Size and electronegativity.
- Deb: And so, do you think one of 'em's going to be more than the other?
- Marie: Yeah. I think... [to Jack] what do you think?
- Jack: I don't know; I don't think it'll be a huge amount more, but I think it'll definitely be different.
- Marie: I think chlorine will be a little bit more than iodide.
- Deb: I was going to say, now which one do you think it's going to be. But okay. [jokingly] Thank you for sharing your opinion with us.
- Jack: No problem.

Marie: Any time.

Kurt: [later, after an intervening episode of non-molecular conversation]
So you think that the chlorine will be more conductive because of...?

Jack: Atomic size.

Marie: The atomic size.

Kurt: But how does that affect... like, why... why are those important?

Jack: So, earlier, we were talking about the H₂O breaking apart the ionic compounds, and how the ionic... that's going to be stronger than the covalent, and that's because of, like... [to Marie]... what did we say?

Marie: The bond energy.

Jack: Bond energy.

Yet again, Kurt suggests – this time explicitly – that the number of dissolved ions in solution relates to conductivity. In asking the remaining students to explain their idea about how size and electronegativity affect conductivity – factors they believe will make it more difficult to disrupt the ionic lattice – Marie, Jack, and Deb again invoke their previous explanation that conductivity relates to the energy associated with dissolution, which they refer to here as “bond energy.”

Although this explanation is incorrect, because the data are consistent with Marie, Jack, and Deb's prediction, it is not explicitly invalidated by the observations of the NaI solution. Each student therefore leaves the laboratory period with experimental validation of their ideas, and we can examine their refined models to discern the degree to which their models reflect the ideas that they expressed in their discussions.

Just like his explanations during the laboratory period, Kurt's refined model is the most distinctly different from the others. He does not relate conductivity to the work involved in breaking apart a crystal lattice, but offers a scientifically correct explanation about the motion of charged particles contributing to conductivity:

“After: When water mixes with the salt, it breaks apart the individual ions:

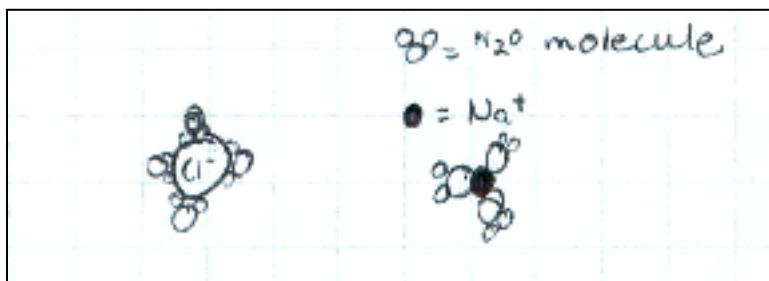


Figure 6.3 Kurt’s molecular-level picture of dissolved salt in his refined model.

These charges [sic] particles in the solution allow electrons to flow, making it highly conductive. When water mixes with the sucrose, it breaks the intermolecular forces holding the sucrose molecules to each other, but does not cause the sucrose to contribute ions to solution. This results in a low conductivity.”

Marie’s final model, however, is consistent with the ideas (which she helped to develop as part of group discussion) about the work involved in separating ions contributing to electrical conductivity. She incorrectly adopts the term “voltage” to describe this, but the ideas are clearly recognizable from the in-class group discussions:

“I predicted that salt and sugar would be arranged in an ionic lattice. This was supported when the dissolution of both salt and sugar produced voltage. Voltage is only produced with the separation of positive and negative charges producing potential electric energy that conducts current and causes a conductive tester to blink.” [...] The separation of charged particles creating voltage is also a result of the water molecules drawing charged particles away from each other. As the water pulls the charges away, voltage is made and the conductivity tester blinks”

She elaborates on this further in her generalization:

“As the positive ions are moved away from the negative ions of the lattice, and vice versa, the distance produces electric potential energy. This is called voltage and is capable of conducting electric current. This is what we measured with the conductivity tester in each instance of dissolved solute in water. Without voltage from separated charged particles, the conductivity would not have been present.”

Deb also helped to develop the ideas about energy associated with dissolution contributing to conductivity. During class discussion, she stated, regarding isopropanol, “I don’t think the H₂O molecules are working as hard to pull it apart,” and she uses similar language in her descriptions of the dissolution of both salt and sugar in her refined model:

“The conductivity of 52 flashes [for NaCl solution] is the measure of how hard the H₂O molecules have to work the break apart the crystal. Based on the experiment ionic compounds have a higher conductivity than covalent compounds.”

“The sugar was added to the water. At this point the polar H₂O molecules began to pull apart the sucrose crystals into sucrose molecules. Due to the covalent bonding of sucrose the H₂O molecules don’t have to work as hard to pull apart the molecules as the ionic bonded salt molecules. Because of less resistance the conductivity of sucrose is lower.”

Jack’s refined model exhibits the shortest and least complex discussion of what contributes to conductivity, consistent with the fact that he participates the least within the in-class discussion on these topics. However, Jack originally contributed an idea

about how, when isopropanol mixed with water, “it’s not really having much of a reaction.” Similar language appears in Jack’s molecular-level model:

“Regardless of where the H₂O attached, the crystals [of salt or sugar] will be broken down one atom at a time and forced to move around in the solution making the visible crystals disappear. In other words, the crystals are broken down into their individual atoms by the strong polar forces of water’s H₂ and O mix. The reaction of the NaCl atom and H₂O atoms in turn creates a big release of energy which leads to a high conductivity rate. NaCl had a class average conductivity rate of 45.3 flashes per 30 seconds. On the other hand, you have sugar and its molecule of C₁₂H₂₂O₁₁ which had a fairly low conductivity of 12 flashes per 30 seconds due to its lack of ionic bonds.”

Here, the actual energy associated with breaking apart the solid is not cited as a reason for conductivity. However, this idea was more fully articulated by Marie and Deb during the class period. Jack’s expressions of similar ideas simply echo theirs, rather than being original contributions on his part; however, Jack’s idea about a chemical reaction being responsible for conductivity does appear in his refined model.

In each case, the ideas developed by students during the laboratory period both echo their initial models, and contribute in substantially to students’ refined models, particularly to their efforts to connect their ideas to experimental evidence. All of the students actively participated in the process of constructing, critiquing, and evaluating each other’s models, and, even though the students did not all arrive at the same model to account for conductivity, all provide clear explanations and mechanisms that link molecular-level behavior to evidence, and reflect their own particular and personal contributions to or perspectives on this discussion.

6.3.2 Case Study 2: “Anyone... Remember?”

The previous case study comes from the laboratory period and student group that, among those analyzed, contained the greatest amount of reflection about molecular-level behavior. We now turn to an example that contains markedly less molecular-level discussion during the laboratory period and correspondingly less detail from students in their written models.

This case study comes from analysis of student video from the first week of the Thermochemistry module that was also examined in conjunction with this study. In this laboratory module, students’ models are centered upon two reactions. The first reaction is an acid/base reaction between dilute aqueous solutions of HCl and NaOH, which results in a modest (8-10 °C) temperature increase for the solutions upon mixing. The second reaction is the combustion reaction between ethanol and oxygen, which is used to heat a sample of water. Students gain experience using standard thermochemical relationships to quantify the amount of energy produced, but the module is intended to prompt students to think about the relationship between chemical bonding and energetics. In the first reaction, chemical bonds are only formed, as H⁺ and OH⁻ ions neutralize and combine to form water molecules, and the reaction releases energy in the form of heat to the surrounding water. In the second reaction, students use an initial input of energy (in the form of a lighted match) to initiate a reaction that involves both breaking and forming of bonds.

For the group studied – one of the five originally constructed video groups, from TA 1’s class – the initial models demonstrate far less molecular-level reflection than those from the previous case study. When molecular-level statements were transcribed, each of the students from the first case study included 4-6 molecular-level statements in

their initial models and 11-20 in their refined models. Students in this group had between 2-6 molecular-level statements in each of their initial models, and between 6-11 molecular-level statements in their refined models. While 48 distinct molecular-level ideas were identified as being expressed by the group in the first case study, this group articulated only 9 molecular-level ideas over the course of the entire laboratory period. Because of the relatively small number of molecular-level discussions during the period, we begin this case study by discussing these, and then turn to students' models:

The most extended episode of molecular-level reflection came during a whole-class discussion. Because the group being recorded was located closest to the blackboard where the instructor was leading this discussion, they are the only members of the class who can be heard participating in the discussion.

Instr: So when we were talking about this system next door, we were talking about the relationship between energy and bonds breaking and forming, what did we say about bonds breaking and forming in this system? [long pause] Anyone... remember? Did we say that bonds were being broken or being formed in this reaction?

Diane: Both.

Instr: Bonds being formed, right? So what did you observe in terms of temperature? [pause] I think everyone saw basically the same thing. Did the temperature increase or decrease?

Ryan: Increase.

Instr: Okay, so the temperature increased, but what is that the temperature of?

Ryan: The water.

Instr: The water. So it's not the system that we're interested in, it's the surrounding right... so the temperature... [writes on board]... when that's happening, like you guys said, when bonds are formed. Does that agree with what we said next door?

Diane: No.
Ryan: Nope.
Instr: When we were over there, we said that when bonds were formed, energy would be...
Cody: Absorbed.
Instr: Absorbed. But here we have data saying that when bonds are formed, energy is released from the system. So something to think about while you're... considering and reflecting and graphing.

In contrast with the instructor's interactions with the laboratory group described in the previous case study, in this excerpt, the instructor (a different individual) provides most of the information, rather than actively soliciting students' ideas. The interactions described below are typical of this group. In one episode, Ryan is responding to a reflection question within the laboratory manual prompting him to draw a molecular-level picture of the acid-base reaction:

Ryan: What are you going to draw?
Cody: Uh, I think for the beginning, like, the.. uh, reactant species, I think it's going to be ionized.
Ryan: Okay. [yawns]

A similar episode occurred in response to a question about the second reaction during the laboratory episode:

Diane: [reading a reflection question from the laboratory manual.] What did this tell us about the combustion reaction?
Cody: That the, uh ethanol gets broken down into its atoms.

Students, however, did not take all of their ideas from the same source; a different student provided ideas regarding the combustion reaction:

Ryan: How are you going to draw a picture of this?

Diane: I drew it like... [Holds up lab manual. Ryan and Cody come over to look. They leave, and then Diane shows lab manual to Tara.]

What is particularly interesting about these students' refined models is that their articulation of the relationship between energetics and bond forming is vague at best. These ideas are only addressed in their discussion in the episode described above, where group members contributed (if minimally) to a whole-class discussion in which the instructor was the primary source of molecular-level ideas.

In her refined model, Diane consistently exhibits a correct conception about the relationship between bonding and energetics, stating in her generalization "When bonds are broken, energy is required and when bonds are formed, energy is released."

However, her model of the acid/base reaction exhibits a misconception about what bonds formed. Diane wrote:

"When the NaOH (aq) and HCl (aq) were combined, bonds were broken, absorbing energy, and then formed again, releasing energy. First, the positive Na ion and the OH anion break apart and the H cation breaks from the Cl anion. Then, the Na cation bonded with the Cl anion and the H cation bonded with the OH anion to form water. Although the system both lost and gained energy, it had an overall loss of energy. This is apparent because the surroundings of the system gained heat, indicating a gain in energy."

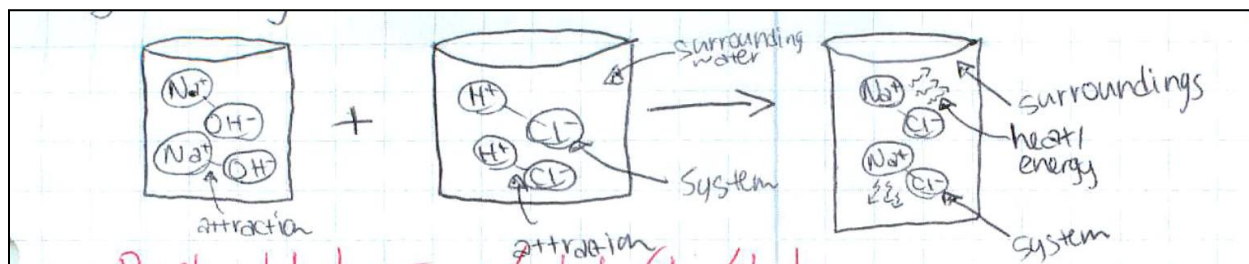


Figure 6.4 Diane's molecular-level picture of the acid/base reaction in her refined model.

Even though the instructor, during the whole class discussion, directly corrected Diane's idea that bonds were both being broken and formed in this reaction, she continued to hold this conception into her refined model, accommodating her theory to account for her observations. The fact that Diane did not personally develop or articulate ideas about bonding during the laboratory period contrasts with the actions of the students in the first case study group, all of whom incorporated their in-class ideas into their refined models to a significant degree.

Aspects of Tara's refined model that related to bond breaking and forming were generally consistent and scientifically correct. She states, about the acid-base reaction:

“When new bonds were formed between $\text{NaCl} + \text{HOH}$, the temperature of the surroundings went up, as well as the energy. The temperature only increased at an instant, because the bonds formed once, instead of continuously forming, making the temperature increase.”

She reasons similarly about the combustion reaction:

“New bonds were formed between the products, $\text{CO}_2 + \text{H}_2\text{O}$, releasing exothermic [sic].”

Her generalization is consistent with both of these statements:

“In general, when bonds are formed, energy is produced and the temperature will increase. If the bonds are formed at an instant, such as mixing, then the temp will increase once and remain constant. If heat is applied and bonds are continuously being formed, then the temperature will continue to rise.”

Cody expressed two conflicting ideas in his model. When talking about the acid-base reaction directly, he correctly relates bond formation with the release of energy:

“Once mixed, the Na^+ and Cl^- stayed in solution, but the H^+ and OH^- combined to form water. Since all the rxns were in solution, no

bonds were broken when they were mixed, so no energy was absorbed. Instead, as the H^+ and the OH^- combined to form H_2O , the formation of bonds released energy which resulted in the temperature increase.”

However, within his generalization, he reverses this relationship, and also seems to conflate thermodynamic systems and surroundings:

“If, during a chemical reaction, bonds are formed, the reaction will be endothermic. If bonds are broken, then the reaction will be exothermic. In our experiment, when the bonds in ethanol broke, the surroundings got warmer, meaning the reaction was exothermic. In the reaction between HCl (aq) and NaOH (aq), the system got warmer, meaning it absorbed heat from the surrounding. During exothermic reactions, the breaking bonds release energy to the surrounding environment. This increases the kinetic energy and temperature of the surrounding.”

Finally, Ryan’s description of the relationship between bond breaking and forming is generally consistent with the incorrect idea that bond breaking releases energy. This is evidenced in his discussion of the reactions:

“On the molecular level when NaOH is mixed with HCl it breaks apart the bonds in each. It forms Na^+ ions and Cl^- bonds [sic] in H_2O . When a bond is broken it gives energy to the surroundings, losing energy. This is seen in the temperature change of the water, the NaOH ’s and HCl ’s surround [sic]. When $\text{CH}_3\text{CH}_2\text{OH}$ is mixed with O_2 gas bonds are broken apart as well. So many bonds are broken that it continually loses energy and transfers it to its surroundings, the water in the can. This can be seen with the temperature increase of the water.”

This incorrect idea is also evidenced in his generalization:

“When bonds are broken, energy is lost, usually in the form of heat. The heat is then transferred to the surroundings of the reaction. When the bonds were broken between NaOH and HCl, the energy as heat is transferred to the matter.”

Although the language in the second portion is somewhat muddled, Ryan indicates a belief that bond breaking is associated with a release of energy, although he is not clear whether it is the system or surroundings that lose energy.

From this case study, we can see that, in a situation where students are not actively and collaboratively working to develop and articulate molecular-level ideas, the ideas that they express in their refined models may not only contradict the evidence that the students collected, but also contradict ideas they were told by their instructor. In such a situation, students may even express contradictory ideas within the same model – a situation analogous to the “context-bound” conceptions of solutions discussed in Chapter 2, and indicative of a shallower understanding of the systems’ molecular-level behavior.

The relationships between student discussion during the laboratory period and the content of post-laboratory refined models cannot be said to be causative based on these limited case studies. However, the pronounced contrast in an entire set of behaviors – the tendency of the instructor in the second case study towards providing information to students versus the instructor in the second case study prompting student elaboration of ideas, students’ proclivity to simply provide each other ideas without discussion in the second case study versus their tendency to reflect on and build off of each other’s ideas in the first and the shallowness of their molecular-level discussions both in class and in their refined models in the second case study versus the

more detailed discussions produced by students in the first case study – speaks toward desirable practices for implementation of inquiry-based laboratory experiments. Specifically, asking questions designed to prompt students to think about molecular-level behaviors, encouraging students to exchange ideas and to check their validity and sensibility both with each other and against experimental evidence, and providing students' with opportunities and encouragement to develop their own understandings (even if these understandings are sometimes incorrect) are associated with a greater depth of reflection. Such discussion during the laboratory period may foster post-laboratory written models that are consistent with experimental evidence and that can be justified with reference to the two key cognitive processes previously identified as being linked with success at transferring knowledge into new contexts.

Chapter 7: Further Research Questions

As described in the National Research Council's 2012 report on Discipline-Based Education Research, "[d]espite its importance in the curriculum, the role of the chemistry laboratory in student learning has gone largely unexamined."⁸³ This research adds to the efforts to ameliorate this deficiency by studying how inquiry-based instructional approaches can be most effectively implemented, and what factors are associated with student engagement in key aspects of cognition in within these approaches.

The questions examined in this dissertation have surrounded the implementation of one particular inquiry-based approach to fostering student learning in the laboratory, but the results of these studies raise an even greater number of further questions that can be addressed in the near-term. In this concluding chapter, I address each chapter's major findings and discuss further potential avenues for investigation stemming from each of them.

In Chapter 2, I outlined a laboratory module that contributed to students' understanding of precipitation reactions, and showed how students' molecular-level ideas within that module (and in other situations) were sensitive to the context in which they were activated. In concert with analysis of the quality of engagement in key thinking processes, the new laboratory model suggests another context and another "transfer question" that can be used to evaluate the depth of students' understandings. Identifying whether students' ideas exhibit similar characteristics to the others described in previous work – and, in particular, whether the ideas students express are linked with key aspects in student models – could help more clearly specify the benefits

of engaging in these key aspects of cognition, as well as their limits. Ultimately, developing and understanding instructional approaches that alleviate the troubling phenomena and inconsistent student ideas described in this chapter is one of the major focuses of this dissertation and related work.

Chapter 3 outlined the effects of introductory activities on students' engagement in several aspects of model construction. Building upon these activities, a second week of activities within the laboratory curriculum provides continued emphasis on instruction in model-writing. To date, the effectiveness of this second week of activities has not been evaluated. At the conclusion of these activities, students rewrite and resubmit the first full refined model that they wrote during the semester, improving their models based upon what they have learned in the activity. The degree of improvement in these resubmissions can be characterized, and also compared to the degree of improvement seen in student resubmissions for similar models from the fall 2008 semester, when students resubmitted models based simply upon instructor feedback.

Another methodology that might be of great interest in elucidating students' thinking processes involved in constructing their written models might be to have students produce their models in more controlled settings, in which they engage in think-aloud processing while writing (or re-writing) their models. Such studies could provide insight into how students consider and incorporate both written feedback and information learned in the process of reviewing one's own or other models.

In Chapter 4, I identified that the accuracy students' self-assessments were associated with the quality of models that they produced. Students' assessments of peers' models were also collected as part of the InSPRE activity, but have not been

analyzed and are a rich data set for potential future analysis. Other research has shown that assessments of performance made by peers are often more accurate than self-assessments.^{62, 84} Several facets of these evaluations can be explored: the raw numerical ratings can allow examination of whether the apparent effects of instructors' stricter or more lenient grading on self-assessments extended to peer assessments as well – and the accuracy of these self-evaluations – may shed light on whether students who assess peers' models more accurately are also more accurate self-evaluators.

In Chapter 5, I examined the relationships between feedback on students' models and their self-assessments. While that chapter included a discussion of potential reasons why divergent feedback may not have resulted in any measurable differences in the quality of models produced, analysis of resubmissions did indicate that students' responses to processed "direct" and "reflective" may be different in some circumstance. This distinction between feedback conditions is therefore ripe for further study. Such contrasts in feedback can be implemented in contexts involving a greater likelihood that students will attend to the feedback, for instance, those involving consecutive activities taking place closer together in time or in which students receive a greater quantity of feedback at a particular time (or a comparable amount of feedback over a shorter period of time). Furthermore, rather than examining the impacts of delivering contrasting styles of feedback – and therefore limiting findings to descriptions of their relative effectiveness – one could analyze the particular types of feedback received by students' who make productive revisions in their resubmissions, who improve the quality of their models over the semester, or who improve their self-evaluation accuracy over the semester, to determine if there were particular sorts of feedback statements associated with these improvements.

As described in Chapter 4, efforts to improve students' self-assessment have historically met with mixed results. One hypothesis not tested is whether regular experience reviewing peers' models and one's own models would improve the accuracy of students' self-assessments or the quality of models they produce, as compared to students who did not participate in regular reviewing activities. Indeed, the development of both the InSPRE activity and the introductory activities involving model-reviewing and critiquing were mutually inspired and informed by each other, and such questions have been investigated in a very general fashion in work surrounding computerized peer-review systems such as CPR (Calibrated Peer Review) ^{47c, 48b-d}. Of particular interest is whether there are aspects of students' writing or reasoning that are more or less amenable to improvement in this way.

In Chapter 6, we described how the instruction and prompts delivered by instructors and via the laboratory manual prompt students' reflection about molecular-level behavior, and how the type and quality of reflection that students exhibit during the laboratory period is echoed in their post-laboratory written models. This work presented only a small cross-section of interactions. To validate and expand this work, additional analysis – both qualitative and quantitative – should be carried out at a variety of educational levels, in a variety of laboratory experiments and inquiry-based frameworks, and in classrooms where there are a variety of instructor personalities and experience levels.

Broader questions. Work towards improving the quality of students' models also has the potential to improve our understanding of the mechanisms by which inquiry-based instruction operates and what aspect of students' cognition is effective at promoting deep understanding. To date, the relationships that have been found between

students' written work and their success at transfer of their understandings into new contexts are correlative in nature. To establish whether these relationships are causative, it will be necessary to create situations in which students' experiences are as comparable as possible, but in which one group of students engages strongly in these key aspects of cognition and another engages more poorly. The design of such an experiment may well rely on knowledge of what instructional factors prompt these aspects of cognition.

Relatively little attention has been paid to individual student differences in this study. Occasional mention has been made of gender or of pre-measures such as performance on a pretest. Factors such as past student experiences with chemistry, attitude and motivation, and epistemologies (beliefs about the nature of knowledge and learning) are likely to have substantial effects on students' engagement in key thinking processes. Such student characteristics can be examined using existing instruments, interview protocols, or naturalistic observations; a plethora of potential relationships and mediating factors exist here, offering a ripe area for further study.

Finally, work must be done to train instructors in the delivery of inquiry-based curricula in all of these aspects – from introducing them, to facilitating in-class discussion, to providing feedback on students' written work. The majority of laboratory course instruction at large research institutions is carried out by graduate teaching assistants, most of whom have very little or no experience with inquiry-based instruction. Helping TAs adjust to teaching using an approach that they themselves are unfamiliar and sometimes uncomfortable with is a substantial challenge, but is an important component of both disseminating curricular innovations and ensuring that they are implemented faithfully and effectively. Due to relatively small populations of

participant TAs, studies that attempt to investigate these questions can be challenging to design and implement.

Concluding remarks. Particularly when instructors are not familiar with inquiry-based instruction (and even when they are), the leap into teaching using such approaches can be large, and there are a multitude of questions to be answered about how to best deliver this instruction. This dissertation has investigated a few of the many instructional and cognitive factors that affect important student outcomes. In the context of one research-based, inquiry-oriented laboratory instructional method, I have identified a variety of factors that impact students' ability to develop high-quality models of chemical systems. The use of reviewing and scaffolding activities as introductions to model-writing, students' personal ability to accurately self-assess the quality of their work, receipt of feedback prompting students to reflect on the quality of their models, and instructor and manual questions explicitly focused at the target areas of cognition are positively associated with higher quality and depth of students' work. This work, we hope, will inform both implementation of these curricula, development and improvement of inquiry-based instruction in the chemistry laboratory more broadly, and contribute to a deeper understanding of the cognitive processes that help foster students' deep understanding of chemistry content.

References: Part I

1. (a) Nurrenbern, S. C.; Pickering, M., Concept-learning versus problem-solving - is there a difference? *Journal of Chemical Education* **1987**, 64 (6), 508-510; (b) Sawrey, B. A., Concept learning versus problem solving: Revisited. *Journal of Chemical Education* **1990**, 67 (3), 253; (c) Nakhleh, M. B., Are our students conceptual thinkers or algorithmic problem solvers? Identifying conceptual students in general chemistry. *Journal of Chemical Education* **1993**, 70 (1), 52; (d) Nakhleh, M. B.; Mitchell, R. C., Concept learning versus problem solving: There is a difference. *Journal of Chemical Education* **1993**, 70 (3), 190; (e) Sanger, M. J.; Greenbowe, T. J., Common student misconceptions in electrochemistry: Galvanic, electrolytic, and concentration cells. *Journal of Research in Science Teaching* **1997**, 34 (4), 377-398.
2. (a) Nakhleh, M. B., Why some students don't learn chemistry - chemical misconceptions. *Journal of Chemical Education* **1992**, 69 (3), 191-196; (b) Calyk, M.; Ayas, A.; Ebenezer, J. V., A review of solution chemistry studies: Insights into students' conceptions. *Journal of Science Education and Technology* **2005**, 14 (1), 29-50; (c) Ozmen, H., Some student misconceptions in chemistry: A literature review of chemical bonding. *Journal of Science Education and Technology* **2004**, 13 (2), 147-159; (d) Garnett, P. J.; Garnett, P. J.; Hackling, M. W., Students' alternative conceptions in chemistry: A review of research and implications for teaching and learning. *Studies in Science Education* **1995**, 25, 69-95.

3. (a) Ault, A., What's Wrong with Cookbooks? *Journal of Chemical Education* **2004**, *81* (11), 1569; (b) Cracolice, M. S.; Monteyne, K., What's wrong with cookbooks? A reply to Ault. *Journal of Chemical Education* **2004**, *81* (11), 1559.
4. Domin, D. S., Considering Laboratory Instruction through Kuhn's View of the Nature of Science. *Journal of Chemical Education* **2009**, *86* (3), 274.
5. (a) Fay, M. E.; Grove, N. P.; Towns, M. H.; Bretz, S. L., A rubric to characterize inquiry in the undergraduate chemistry laboratory. *Chemistry Education Research and Practice* **2007**, *8* (2), 212-219; (b) Bruck, L. B.; Bretz, S. L.; Towns, M. H., Characterizing the level of inquiry in the undergraduate laboratory. *Journal of College Science Teaching* **2008**, *38* (1), 52-58; (c) Pavelich, M. J.; Abraham, M. R., An inquiry format laboratory program for general chemistry. *Journal of Chemical Education* **1979**, *56* (2), 100-103; (d) Domin, D. S., A review of laboratory instruction styles. *Journal of Chemical Education* **1999**, *76* (4), 543.
6. Piaget, J.; Elkind, D.; Tenzer, A., *Six psychological studies*. Random House New York: 1967.
7. (a) Bodner, G. M., Constructivism: A theory of knowledge. *Journal of Chemical Education* **1986**, *63* (10), 873-877; (b) Bodner, G.; Klobuchar, M.; Geelan, D., The many forms of constructivism. *Journal of Chemical Education* **2001**, *78* (8), 1107.
8. Staver, J. R., Constructivism: Sound theory for explicating the practice of science and science teaching. *Journal of Research in Science Teaching* **1998**, *35* (5), 501-520.

9. Rickey, D.; Teichert, M. A.; Tien, L. T., Model-Observe-Reflect-Explain (MORE) Thinking Frame Instructions: Promoting Reflective Laboratory Experiences to Improve Understanding of Chemistry. In *Chemists' Guide to Effective Teaching, Volume I*, Pienta, N. J.; Cooper, M. M.; Greenbowe, T. J., Eds. Pearson Prentice-Hall: Upper Saddle River, NJ, 2009; pp 158-172.
10. (a) Tien, L. T. Fostering Expert Inquiry Skills and Beliefs about Chemistry through the MORE Laboratory Experience. University of California, Berkeley, CA., Ph.D. Thesis, 1998; (b) Rickey, D. The Effects of Laboratory Curriculum and Instruction on Undergraduate Students' Understanding of Chemistry. University of California, Berkeley, CA, Ph.D. Thesis, 1999; (c) Tien, L. T.; Rickey, D.; Stacy, A. M., The MORE Thinking Frame: Guiding students' thinking in the laboratory. *Journal of College Science Teaching* **1999**, *28*, 318-324.
11. (a) Whitehead, A. N., *The aims of education*. New York: Macmillan: 1929; (b) Halpern, D. F.; Hakel, M. D., Applying the science of learning to the university and beyond: Teaching for long-term retention and transfer. *Change: The Magazine of Higher Learning* **2003**, *35* (4), 36-41; (c) Rickey, D.; Stacy, A. M., The role of metacognition in learning chemistry. *Journal of Chemical Education* **2000**, *77* (7), 915-920; (d) Spencer, J. N., New directions in teaching chemistry: A philosophical and pedagogical basis. *Journal of Chemical Education* **1999**, *76* (4), 566.
12. Tien, L. T.; Teichert, M. A.; Rickey, D., Effectiveness of a MORE laboratory module in prompting students to revise their molecular-level ideas about solutions. *Journal of Chemical Education* **2007**, *84* (1), 175-181.

13. Teichert, M. A.; Tien, L. T.; Dysleski, L.; Rickey, D., *Cognition and Instruction*, under review.
14. (a) Silberberg, M., *Chemistry: The Molecular Nature of Matter and Change*. McGraw-Hill Education: 2011; (b) Brown, T. L.; Brown, T. E.; LeMay, H. E.; Bursten, B. E.; Murphy, C.; Woodward, P., *Chemistry: The Central Science*. Pearson Prentice Hall: 2012; (c) Chang, R.; Goldsby, K., *Chemistry*. McGraw-Hill Education: 2012.
15. (a) Ozmen, H.; Alipaşa, A., Students' difficulties in understanding of the conservation of matter in open and closed-system chemical reactions. *Chemistry Education Research and Practice* **2003**, 4 (3), 279-290; (b) Kelly, R. M.; Barrera, J. H.; Mohamed, S. C., An Analysis of Undergraduate General Chemistry Students' Misconceptions of the Submicroscopic Level of Precipitation Reactions. *Journal of Chemical Education* **2010**, 87 (1), 113-118.
16. Furlong, W. R.; Rubinski, M. A.; Indralingam, R., The Method of Continuous Variation: A Laboratory Investigation of the Formula of a Precipitate. *Journal of Chemical Education* **2013**.
17. Maines, L. L.; Bruch, M. D., Identification of Unknown Chloride Salts Using a Combination of Qualitative Analysis and Titration with Silver Nitrate: A General Chemistry Laboratory. *Journal of Chemical Education* **2012**, 89 (7), 933-935.
18. Bonomo, R. P.; Tabbi, G.; Vagliasindi, L. I., Solubility and Solubility Product Determination of a Sparingly Soluble Salt: A First-Level Laboratory Experiment. *Journal of Chemical Education* **2011**, 89 (4), 545-547.

19. DeMeo, S., Using limiting-excess stoichiometry to introduce equilibrium calculations: A discrepant event laboratory activity involving precipitation reactions. *Journal of Chemical Education* **2002**, 79 (4), 474-475.
20. Gunn, E., Conductivity tests for reactions which go to an end. *Journal of Chemical Education* **1940**, 17 (8), 385-386.
21. (a) Marcum-Dietrich, N. I.; Ford, D. J., The place for the computer is in the laboratory: An investigation of the effect of computer probeware on student learning. *Journal of Computers in Mathematics and Science Teaching* **2002**, 21 (4), 361-379; (b) Brown, M., Streamlining Science: Three New Science Tools Make Data Collection a Snap. *Technology & Learning* **2006**, 26 (6), 16.
22. (a) Newton, T. A.; Hill, B. A., Using conductivity devices in nonaqueous solutions I: Demonstrating the S_N1 mechanism. *Journal of Chemical Education* **2004**, 81 (1), 58-60; (b) Ratliff, J. L., Chemistry for the visually impaired. *Journal of Chemical Education* **1997**, 74 (6), 710-711; (c) Rettich, T. R.; Battino, R., An inexpensive and easily constructed device for quantitative conductivity experiments. *Journal of Chemical Education* **1989**, 66 (2), 168; (d) Zawacky, S. K. S., A cheap, semiquantitative hand-held conductivity tester. *Journal of Chemical Education* **1995**, 72 (8), 728-729.
23. (a) Holdsworth, D. K., Conductivity titrations-A microcomputer approach. *Journal of Chemical Education* **1986**, 63 (1), 73-74; (b) Smith, K. C.; Edionwe, E.; Michel, B., Conductimetric titrations: A Predict-Observe-Explain activity for general chemistry. *Journal of Chemical Education* **2010**, 87 (11), 1217-1221.

24. Sanger, M. J.; Danner, M., Aqueous ammonia or ammonium hydroxide? Identifying a base as strong or weak. *Journal of Chemical Education* **2010**, *87* (11), 1213-1216.
25. Dominguez, A.; Fernandez, A.; Gonzalez, N.; Iglesias, E.; Montenegro, L., Determination of critical micelle concentration of some surfactants by three techniques. *Journal of Chemical Education* **1997**, *74* (10), 1227-1231.
26. (a) Chesick, J.; Patterson Jr, A., Determination of reaction rates with an AC conductivity bridge: A student experiment. *Journal of Chemical Education* **1960**, *37* (5), 242-244; (b) Marzik, J. V.; Sabatelli, A. D.; Fitzgerald, P. J.; Sarneski, J. E., A study of the kinetics of the substitution reaction of square-planar platinum(ii) complexes as monitored by conductance measurements. *Journal of Chemical Education* **1981**, *58* (7), 589-592; (c) Newton, T. A.; Hill, B. A., Using conductivity devices in nonaqueous solutions II: Demonstrating the S_N2 mechanism. *Journal of Chemical Education* **2004**, *81* (1), 61-62; (d) Pace, R. D.; Regmi, Y., The Finkelstein reaction: Quantitative reaction kinetics of an S_N2 reaction using nonaqueous conductivity. *Journal of Chemical Education* **2006**, *83* (9), 1344-1348.
27. Kooser, A. S.; Jenkins, J. L.; Welch, L. E., Acid-base indicators: A new look at an old topic. *Journal of Chemical Education* **2001**, *78* (11), 1504-1506.
28. Lucasse, W. W.; Abrahams, H. J., Conductance titrations in non-aqueous solvents and their use as a pedagogical aid. *Journal of Chemical Education* **1930**, *7* (2), 341-348.
29. Mattox, A. C.; Reisner, B. A.; Rickey, D., What happens when chemical compounds are added to water? An introduction to the Model-Observe-Reflect-

- Explain (MORE) Thinking Frame. *Journal of Chemical Education* **2006**, 83 (4), 622-624.
30. Vernier Logger Pro. <http://www.vernier.com/products/software/lp/> (accessed July 9, 2013).
31. Rosenthal, L. C.; Nathan, L. C., A conductimetric-potentiometric titration for an advanced laboratory. *Journal of Chemical Education* **1981**, 58 (8), 656.
32. Blair, A. C.; Fisher, E. R.; Rickey, D., Discovering nanoscience. *Science* **2012**, 337, 1056-1057.
33. Landis, J. R.; Koch, G. G., The measurement of observer agreement for categorical data. *Biometrics* **1977**, 159-174.
34. Teichert, M. A.; Tien, L. T.; Anthony, S.; Rickey, D., Effects of context on students' molecular-level ideas. *Int. J. Sci. Educ.* **2008**, 30 (8), 1095-1114.
35. (a) DiSessa, A., Knowledge in pieces. In *Constructivism in the Computer Age*, Forman, G.; Pufall, P., Eds. Erlbaum: Hillsdale, NJ, 1988; pp 49-70; (b) DiSessa, A. A., Toward an epistemology of physics. *Cognition and instruction* **1993**, 10 (2-3), 105-225; (c) DiSessa, A. A.; Gillespie, N. M.; Esterly, J. B., Coherence versus fragmentation in the development of the concept of force. *Cognitive Science* **2004**, 28 (6), 843-900.
36. Özdemir, G.; Clark, D. B., An overview of conceptual change theories. *Eurasia Journal of Mathematics, Science & Technology Education* **2007**, 3 (4), 351-361.
37. (a) Posner, G. J.; Strike, K. A.; Hewson, P. W.; Gertzog, W. A., Accommodation of a scientific conception: Toward a theory of conceptual change. *Science education* **1982**, 66 (2), 211-227; (b) Ioannides, C.; Vosniadou, S., Exploring the changing meanings of force: From coherence to fragmentation. *Cognitive Science*

- Quarterly* **2002**, 2 (1), 5-61; (c) Chi, M. T., Commonsense conceptions of emergent processes: Why some misconceptions are robust. *The journal of the learning sciences* **2005**, 14 (2), 161-199.
38. (a) Council, N. R., *National Science Education Standards*. National Academy Press: Washington, DC, 1996; (b) Rutherford, F. J.; Ahlgren, A., *Science for All Americans*. American Association for the Advancement of Science: Washington, DC, 1990; (c) Society, A. C., *Undergraduate Professional Education in Chemistry: ACS Guidelines and Evaluations Procedures for Bachelor's Degree Programs*. American Chemical Society: Washington, DC, 2008.
39. (a) Cooper, M. M., Cooperative chemistry laboratories. *Journal of Chemical Education* **1994**, 71 (4), 307; (b) Wink, D. J.; Gislason, S. F.; Keuhn, J. E., *Working with Chemistry: A Laboratory Inquiry Program*. 2nd edition ed.; W.H. Freeman: New York, NY, 2005; (c) Bauer, R. D.; Birk, J. P.; Sawyer, D. J., *Laboratory Inquiry in Chemistry*. 3rd edition ed.; Brooks/Cole: Belmont, CA, 2009.
40. Greenbowe, T. J.; Hand, B. M., Introduction to the Science Writing Heuristic. In *Chemists' Guide to Effective Teaching*, Pienta, N. J.; Cooper, M. M.; Greenbowe, T. J., Eds. Pearson Prentice Hall: Upper Saddle River, NJ, 2005; pp 140-154.
41. (a) Rudd, J. A.; Greenbowe, T. J.; Hand, B., Recrafting the general chemistry laboratory report. *Journal of College Science Teaching* **2002**, 31 (4), 230-34; (b) Pooch, J. R.; Burke, K.; Greenbowe, T. J.; Hand, B., Using the Science Writing Heuristic in the general chemistry laboratory to improve students' academic performance. *Journal of Chemical Education* **2007**, 84 (8), 1371-1379; (c) Rickey, D.; Tien, L. T., Research on Learning in the Chemistry Laboratory: A

- Trajectory Connecting Student Outcomes to Thinking Processes. In *Trajectories of Chemistry Education Innovation*, Holme, T. A.; Cooper, M. M. V.-N., P., Eds. American Chemical Society: Washington, DC., Invited Contribution.
42. (a) Kelly, G. J.; Takao, A., Epistemic levels in argument: An analysis of university oceanography students' use of evidence in writing. *Science education* **2002**, *86* (3), 314-342; (b) Sampson, V.; Walker, J. P., Argument-driven inquiry as a way to help undergraduate students write to learn by learning to write in chemistry. *Int. J. Sci. Educ.* **2012**, *34* (10), 1443-1485.
43. (a) Modell, H. I., Preparing students to participate in an active learning environment. *Advances in Physiology Education.* **1996**, *270*, 69-77; (b) Abd-El-Khalick, F., Developing deeper understandings of nature of science: The impact of a philosophy of science course on preservice science teachers,Â views and instructional planning. *Int. J. Sci. Educ.* **2005**, *27* (1), 15-42.
44. (a) Aydin, S.; Aydemir, N.; Boz, Y.; Cetin-Dindar, A.; Bektas, O., The contribution of constructivist instruction accompanied by concept mapping in enhancing pre-service chemistry teachers' conceptual understanding of chemistry in the laboratory course. *Journal of Science Education and Technology* **2009**, *18* (6), 518-534; (b) Stains, M.; Escriu-Sune, M.; Molina Alvarez de Santizo, M. L.; Sevian, H., Assessing secondary and college students' implicit assumptions about the particulate nature of matter: development and validation of the structure and motion of matter survey. *Journal of Chemical Education* **2011**, *88* (10), 1359-1365.
45. (a) Bruck, L. B.; Towns, M. H., Preparing students to benefit from inquiry-based activities in the chemistry laboratory: guidelines and suggestions. *Journal of*

- Chemical Education* **2009**, 86 (7), 820-822; (b) Lewis, J., The Effectiveness of Mini-Projects as a Preparation for Open-Ended Investigations. In *Teaching and Learning in the Science Laboratory*, Psillos, D.; Niedderer, H., Eds. Kluwer Academic Publishers: : Norwell, MA, 2002.
46. (a) Burke, K.; Greenbowe, T. J.; Hand, B. M., Implementing the science writing heuristic in the chemistry laboratory. *Journal of Chemical Education* **2006**, 83 (7), 1032-1038; (b) Green, W. J.; Elliott, C.; Cummins, R. H., "Prompted" inquiry-based learning in the introductory chemistry laboratory. *Journal of Chemical Education* **2004**, 81 (2), 239-241; (c) Cacciatore, K. L.; Sevian, H., Teaching lab report writing through inquiry: a green chemistry stoichiometry experiment for General Chemistry. *Journal of Chemical Education* **2006**, 83 (7), 1039-1041.
47. (a) McNeill, K. L.; Krajcik, J., Scientific explanations: Characterizing and evaluating the effects of teachers' instructional practices on student learning. *Journal of Research in Science Teaching* **2008**, 45 (1), 53-78; (b) McDonald, C. V., The influence of explicit nature of science and argumentation instruction on preservice primary teachers' views of nature of science. *Journal of Research in Science Teaching* **2010**, 47 (9), 1137-1164; (c) Margerum, L. D.; Gulsrud, M.; Manlapez, R.; Rebong, R.; Love, A., Application of calibrated peer review (CPR) writing assignments to enhance experiments with an environmental chemistry focus. *Journal of Chemical Education* **2007**, 84 (2), 292-295; (d) Gerdeman, R. D.; Russell, A. A.; Worden, K. J., Web-Based Student Writing and Reviewing in a Large Biology Lecture Course. *Journal of College Science Teaching* **2007**, 36 (5), 46-52.

48. (a) Pelaez, N. J., Problem-based writing with peer review improves academic performance in physiology. *Advances in Physiology Education* **2002**, *26* (3), 174-184; (b) Hartberg, Y.; Gunersel, A. B.; Simpsom, N. J.; Balester, V., Development of Student Writing in Biochemistry Using Calibrated Peer Review. *Journal of the Scholarship of Teaching and Learning* **2008**, *2* (1), 29-44; (c) Gunersel, A. B.; Simpson, N. J.; Aufderheide, K. J.; Wang, L., Effectiveness of Calibrated Peer Review [TM] for Improving Writing and Critical Thinking Skills in Biology Undergraduate Students. *Journal of the Scholarship of Teaching and Learning* **2008**, *8* (2), 25-37; (d) Walvoord, M. E.; Hoefnagels, M. H.; Gaffin, D. D.; Chumchal, M. M.; Long, D. A., An analysis of calibrated peer review (CPR) in a science lecture classroom. *Journal of College Science Teaching* **2008**, *37* (4), 66.
49. (a) Birk, J. P.; Lawson, A. E., The persistence of the candle-and-cylinder misconception. *Journal of Chemical Education* **1999**, *76* (7), 914-916; (b) Krnel, D.; Glazar, S. A., "Experiment with a Candle" without a candle. *Journal of Chemical Education* **2001**, *78* (7), 914.
50. Walczyk, J. J.; Hall, V. C., Effects of examples and embedded questions on the accuracy of comprehension self-assessments. *Journal of Educational Psychology* **1989**, *81* (3), 435-437.
51. Hacker, D. J.; Bol, L.; Horgan, D. D.; Rakow, E. A., Test prediction and performance in a classroom context. *Journal of Educational Psychology* **2000**, *92* (1), 160-170.

52. Sluijsmans, D.; Dochy, F.; Moerkerke, G., Creating a learning environment by using self-, peer-and co-assessment. *Learning Environments Research* **1998**, *1* (3), 293-319.
53. Everson, H. T.; Tobias, S., The ability to estimate knowledge and performance in college: A metacognitive analysis. *Instructional Science* **1998**, *26* (1), 65-79.
54. Potgieter, M.; Ackermann, M.; Fletcher, L., Inaccuracy of self-evaluation as additional variable for prediction of students at risk of failing first-year chemistry. *Chem. Educ. Res. Pract.* **2010**, *11* (1), 17-24.
55. Bol, L.; Hacker, D. J.; O'Shea, P.; Allen, D., The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *Journal of Experimental Education* **2005**, *73* (4), 269-290.
56. Kelemen, W. L.; Winningham, R. G.; Weaver III, C. A., Repeated testing sessions and scholastic aptitude in college students' metacognitive accuracy. *European Journal of Cognitive Psychology* **2007**, *19* (4-5), 689-717.
57. Labuhn, A. S.; Zimmerman, B. J.; Hasselhorn, M., Enhancing students' self-regulation and mathematics performance: the influence of feedback and self-evaluative standards. *Metacognition and Learning* **2010**, *5* (2), 173-194.
58. Miller, T. M.; Geraci, L., Unskilled but aware: Reinterpreting overconfidence in low-performing students. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **2011**, *37* (2), 502-506.
59. (a) Thiede, K. W.; Anderson, M.; Therriault, D., Accuracy of metacognitive monitoring affects learning of texts. *Journal of educational psychology* **2003**, *95* (1), 66; (b) Thiede, K. W.; Dunlosky, J.; Griffin, T. D.; Wiley, J., Understanding the delayed-keyword effect on metacomprehension accuracy.

Journal of Experimental Psychology: Learning, Memory, and Cognition **2005**, 31 (6), 1267-1280.

60. (a) Vadhan, V.; Stander, P., Metacognitive ability and test performance among college students. *The Journal of Psychology* **1994**, 128 (3), 307-309; (b) Bol, L.; Hacker, D. J., A comparison of the effects of practice tests and traditional review on performance and calibration. *The Journal of experimental education* **2001**, 69 (2), 133-151; (c) Nietfeld, J. L.; Cao, L.; Osborne, J. W., Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *The Journal of Experimental Educational* **2005**, 7-28; (d) Nietfeld, J. L.; Cao, L.; Osborne, J. W., The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacognition and Learning* **2006**, 1 (2), 159-179; (e) Burns, D. J., An examination of the accuracy of students' expected grade. *Academy of Educational Leadership Journal* **2007**, 11 (3), 45-58; (f) Burns, D. J., Will I do as well on the final exam as I expect? An examination of students' expectations. *Journal of the Scholarship of Teaching and Learning* **2008**, 8 (3), 1-19; (g) Hacker, D. J.; Bol, L.; Bahbahani, K., Explaining calibration accuracy in classroom contexts: the effects of incentives, reflection, and explanatory style. *Metacognition and Learning* **2008**, 3 (2), 101-121; (h) Miller, T. M.; Geraci, L., Training metacognition in the classroom: the influence of incentives and feedback on exam predictions. *Metacognition and learning* **2011**, 6 (3), 303-314; (i) Thompson, D. R., Promoting Metacognitive Skills in Intermediate Spanish: Report of a Classroom Research Project. *Foreign Language Annals* **2012**, 45 (3), 447-462; (j) Bol, L.; Hacker, D. J.; Walck, C. C.; Nunnery, J. A., The effects of individual or group guidelines on the calibration

- accuracy and achievement of high school biology students. *Contemporary Educational Psychology* **2012**, 37 (4), 280-287; (k) Was, C. A.; Beziat, T. L.; Isaacson, R. M., Improving Metacognition in a College Classroom: Does Enough Practice Work? *Journal of Research in Education* **2013**, 23 (1), 77-93; (l) Barenberg, J.; Dutke, S., Metacognitive monitoring in university classes: anticipating a graded vs. a pass-fail test affects monitoring accuracy. *Metacognition and Learning* **2013**, 1-23.
61. Liang, J. C.; Tsai, C. C., Learning through science writing via online peer assessment in a college biology course. *The Internet and Higher Education* **2010**, 13 (4), 242-247.
62. Helzer, E. G.; Dunning, D., Why and when peer prediction is superior to self-prediction: The weight given to future aspiration versus past achievement. *Journal of Personality and Social Psychology* **2012**, 103 (1), 38-53.
63. Papinczak, T.; Young, L.; Groves, M.; Haynes, M., An analysis of peer, self, and tutor assessment in problem-based learning tutorials. *Medical teacher* **2007**, 29 (5), 122-132.
64. Stacy, A. M., *Living by Chemistry*. W.H. Freeman: 2012.
65. Robinson, R., Calibrated Peer Review™: An Application To Increase Student Reading & Writing Skills. *The American Biology Teacher* **2001**, 63 (7), 474-480.
66. Cohen, J., *Statistical power analysis for the behavioral sciences*. Routledge Academic: 1988.
67. Schraw, G., A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning* **2009**, 4 (1), 33-45.

68. (a) Lichtenstein, S.; Fischhoff, B., Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance* **1977**, *20* (2), 159-183; (b) Krug, K., The relationship between confidence and accuracy: Current thoughts of the literature and a new area of research. *Applied Psychology in Criminal Justice* **2007**, *3* (1), 7-41; (c) Galotti, K. M., *Cognitive psychology in and out of the laboratory*. SAGE Publications, Incorporated: 2013.
69. Dunlosky, J.; Rawson, K. A., Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction* **2012**, *22* (4), 271-280.
70. Maki, R. H.; Shields, M.; Wheeler, A. E.; Zacchilli, T. L., Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology* **2005**, *97* (4), 723.
71. (a) Kluger, A. N.; DeNisi, A., The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin* **1996**, *119* (2), 254-284; (b) Hattie, J.; Timperley, H., The power of feedback. *Review of Educational Research* **2007**, *77* (1), 81-112.
72. Butler, D. L.; Winne, P. H., Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research* **1995**, *65* (3), 245-281.
73. Paris, S. G.; Winograd, P., Promoting metacognition and motivation of exceptional children. *Remedial and Special Education* **1990**, *11* (6), 7-15.
74. Nelson, T. O., Consciousness and metacognition. *American psychologist* **1996**, *51* (2), 102.

75. (a) Sadler, D. R., Formative assessment and the design of instructional systems. *Instructional science* **1989**, *18* (2), 119-144; (b) Gibbs, G.; Simpson, C., Conditions under which assessment supports students' learning. *Learning and teaching in higher education* **2004**, *1* (1), 3-31; (c) Nicol, D. J.; Macfarlane-Dick, D., Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education* **2006**, *31* (2), 199-218; (d) Clark, I., Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review* **2012**, *24* (2), 205-249.
76. Sadler, D. R., Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education* **2010**, *35* (5), 535-550.
77. Glenberg, A. M.; Sanocki, T.; Epstein, W.; Morris, C., Enhancing calibration of comprehension. *Journal of Experimental Psychology-General* **1987**, *116* (2), 119-136.
78. Montgomery, J. L.; Baker, W., Teacher-written feedback: Student perceptions, teacher self-assessment, and actual teacher performance. *Journal of Second Language Writing* **2007**, *16* (2), 82-99.
79. Chi, M. T., Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science* **2009**, *1* (1), 73-105.
80. Smith-Lovin, L.; Brody, C., Interruptions in group discussions: The effects of gender and group composition. *American Sociological Review* **1989**, 424-435.
81. Barab, S. A.; Hay, K. E.; Yamagata-Lynch, L. C., Constructing networks of action-relevant episodes: An in situ research methodology. *The Journal of the Learning Sciences* **2001**, *10* (1-2), 63-112.

82. Bruck, L. B.; Towns, M.; Bretz, S. L., Faculty perspectives of undergraduate chemistry laboratory: Goals and obstacles to success. *Journal of Chemical Education* **2010**, *87* (12), 1416-1424.
83. Singer, S. R.; Nielsen, N. R.; Schweingruber, H. A., *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*. National Academies Press: 2012.
84. (a) Dunning, D.; Heath, C.; Suls, J. M., Flawed self-assessment implications for health, education, and the workplace. *Psychological science in the public interest* **2004**, *5* (3), 69-106; (b) Sullivan, M. E.; Hitchcock, M. A.; Dunnington, G. L., Peer and self assessment during problem-based tutorials. *The American Journal of Surgery* **1999**, *177* (3), 266-269.

PART II

SOLVATION IN SUPERCRITICAL CARBON DIOXIDE/ETHANOL MIXTURES
STUDIED BY MOLECULAR DYNAMICS SIMULATION

Chapter 8: Introduction: Solvation in Supercritical CO₂/Cosolvent Systems

The role of solvents in chemical systems is difficult to understate. The fact that one substance can be dissolved in another allows the molecules or ions of the dissolved solute to be mobile, facilitating processes such as chemical reactions or separations. As described in any general chemistry course, not every substance is soluble in every solvent; the nature of the interactions between solute and solvent species facilitate or hinder dissolution.

Such phenomena are typically considered with respect to the liquid phase, where intermolecular attractions are of sufficient strength to hold component molecules or ions in close proximity to each other, but are weak enough to allow chemical species to readily move past each other. However, the role of attractive interactions in facilitating solvation and governing the properties of such systems is different in the supercritical regime, where fluids are held at sufficiently high temperatures that intermolecular attractions are overcome by the kinetic energy of component molecules, but sufficiently high pressures that the densities (and therefore the proximity of molecules) are comparable to the liquid phase.

In 1879, Hannay and Hogarth¹ described how a number of substances – including ionic compounds such as metal chlorides and iodides, and organic molecules such as chlorophyll and aniline dyes – were insoluble in liquid solvents (such as short-chain alcohols, ether, or carbon tetrachloride) but actually became soluble in the same substance at conditions beyond the critical point. Interest in supercritical solvents has grown markedly in recent decades, and among the most widely used is supercritical

carbon dioxide (sc-CO₂), which has proven to be a versatile solvent for separation processes ranging from the extraction of nicotine from tobacco,² caffeine from coffee beans,³ and asphaltenes from crude oil.⁴

Compared to traditional organic solvents, sc-CO₂ offers a range of advantages: it is easy to purify, generally nonreactive, and non-toxic to humans. Additionally, it is easy and economical to regenerate – following extraction at high pressure in fluid sc-CO₂, the pressure can be reduced, precipitating any solute and allowing gaseous CO₂ to be recovered in an energy-efficient manner.^{3,5} CO₂'s critical point, at 7.38 MPa (or 72.8 atm) and 304.2 K (31.1°C or 88.0°F), is easily accessible under laboratory or industrial conditions. It is the accessibility of the supercritical regime that contributes to another benefit of sc-CO₂ – the potential for control over chemical processes. Within the supercritical regime, a great deal of thermodynamic control can be exerted due to the fact that substantial changes in density arise with relatively small changes within this relatively accessible temperature and pressure range. This results in similarly large variations in density-dependent properties, such as viscosity, dielectric constant, and, for solvated species, solute mobility and solubility.⁶

However, despite the desirable characteristics of sc-CO₂, attractive interactions still do facilitate the dissolving of substances in supercritical fluids, and carbon dioxide is somewhat problematic in this regard. Carbon dioxide is essentially non-polar, and therefore possesses only very weak attractive interactions (and correspondingly low solubilities) with the polar organic solutes that are often of great interest in industrial and analytical processes.

This difficulty has been ameliorated by the introduction of cosolvents, sometimes called *entrainers* – typically small polar molecules such as short-chain alcohols. Since

the early 1980's, researchers have observed enhanced solubility of some polar organic compounds within binary supercritical fluids as compared to pure supercritical fluids.⁵ For example, the addition of 6.5% mole percent ethanol to supercritical CO₂ at about 318.5 K and 100 bar increases the solubility of salicylic acid by two orders of magnitude.⁸

Solvation of compounds within supercritical fluids is believed to be associated with clustering of solvent molecules around a solute. Solvatochromic experiments, measuring the influence of solvent molecules on the spectroscopic properties of solute molecules, have confirmed local density enhancements of polar cosolvent around solute molecules^{6b} and the local mole fraction of the polar cosolvent around the solute is enhanced compared to the bulk composition.⁹

Ting et al,¹⁰ in modeling supercritical-fluid cosolvent systems, have found that solubility increases as an exponential function of cosolvent concentration within binary sc mixtures, in contrast to liquid mixtures, where relationships between solubility and cosolvent concentration tend to be more linear. These findings demonstrate that unique chemical behaviors within the supercritical regime – possibly related to local inhomogeneities – occur within these systems and are worthy of further description and study. Understanding the nature of the interactions between two components of a binary mixture within the supercritical regime is also beneficial if one wishes to control processes that occur within these fluids.

In this study, we focus in particular on the CO₂/ethanol system, as the strong intermolecular interaction of hydrogen bonding is likely to play a significant role within these systems and is both amenable to study by simulation and experiment. Although this study strictly focuses on simulations of these systems, the potential to make contact

with experimental data is crucial in order to ultimately validate computational or theoretical models.

8.1 Previous Studies of sc-CO₂/Alcohol Systems

Our review here of past studies – both experiments and simulation – on the structure and dynamics of sc-CO₂/ethanol and similar mixtures – will encompass mixtures of CO₂ and small chain alcohols (methanol through *n*-hexanol) at temperatures near or above the critical temperature of CO₂. Studies of CO₂-expanded liquids (generally, liquid mixtures at or above 300K and containing large mole fractions of the alcohol), as related phenomena may occur below the critical temperature, and as it is possible to vary thermodynamic conditions (including temperature, pressure, and CO₂ mole fraction) in such a manner as to also vary chemical and physical properties continuously from those of neat organic solvents.^{6a, 11}

Experimentally, two classes of experiments – IR and NMR – have elucidated the structure of sc-CO₂/alcohol mixtures. IR experiments typically probe the vibrational frequency of the O-H stretch within alcohols; NMR experiments most often probe the ¹H signal from the hydroxyl proton – and both of these signals shift systematically when the relevant alcohol molecule is hydrogen bonded. From the degree of shift, the relative fraction of monomeric and hydrogen-bonded alcohols within these mixtures can be determined.

Two types of computational methods have also contributed to our understanding of these mixtures. Monte Carlo algorithms, which generate a set of molecular configurations based on sampling pseudorandomly (using the Boltzmann distribution for NVT simulations, for instance) to simulate sampling from the canonical ensemble,

enable examination of structural properties, but not dynamics. Molecular dynamics simulations, by contrast, allow researchers to approximate the dynamical behavior of molecules by simulating them as interacting particles and then applying Newton's equations of motion in discrete timesteps, and monitoring behavior as a function of time. With this type of simulation data, one can examine dynamical properties such as diffusion coefficients, hydrogen bond lifetimes, or the reorganization rate of the first solvation shell.

IR Experiments

To date, IR experiments have yielded relatively limited information about the structure of sc-CO₂/alcohol mixtures. Fulton, Yee, and Smith¹² determined, based on the intensity of IR peaks associated with monomeric and oligomeric ethanol molecules, that the mixture could be effectively modeled as containing an increasing quantity of hydrogen-bonded (non-monomeric) ethanol molecules as the ethanol mole fraction within the mixture increases. Unexpectedly high concentrations of the ethanol monomer within sc-CO₂ (as compared with other solvents), however, caused Fulton et al¹² to suggest the existence of a weak but favorable interaction between CO₂ and ethanol.

Lalanne et al¹³ used frequency shifts of the O-H stretch to estimate the local density around dilute ethanol. Because of the low ethanol concentrations in their experiments, they were unable to probe hydrogen-bonding phenomena, but they did reported a density enhancement of CO₂ around ethanol of up to 70 percent greater than bulk density, and concluded that a favorable CO₂-ethanol interaction also exists.

NMR Experiments

A similar local density enhancement for dilute ethanol within CO₂ can be deduced from NMR measurements, wherein the hydroxyl proton signal is shifted to a higher frequency with increasing local density. Kanabuko et al¹⁴ were able to estimate this local density enhancement for dilute ethanol within CO₂; Lalanne et. al¹³ compared this data to data obtained by IR studies and found a greater local density enhancement in IR studies. They attributed the greater local density enhancement found in IR studies to the IR experiment probing a narrower region, as the vibrational phenomena are dominated by short-range dispersion interactions, whereas the quadrupolar field of CO₂ that influences NMR measurements acts over a longer range.

NMR experiments in more concentrated solutions of alcohols within CO₂ also confirm the existence of hydrogen bonding within these systems, again by the shifts of the hydroxyl proton. Schnabel, Srivastava, Vrabec, and Hasse¹⁵ extended these results by directly comparing MD simulations of methanol in supercritical CO₂ with NMR data. Modeling the chemical shift as a combination of contributions from methanol monomers, hydrogen-bond donors, and acceptors, they were able to reproduce experimental ¹H shifts, confirming a relatively modest trend of fewer monomers and more aggregates at lower temperatures and a more pronounced trend of fewer monomers and more aggregates at higher methanol compositions.

Finally, ¹³C spin-relaxation studies carried out by Taylor, Bai, Mayne, and Grant¹⁶ concluded that methanol in supercritical CO₂ self-associates, and that the fraction of methanol molecules remaining as a monomer decreases as the fluid density increases.

Quantum Calculations

Experimental studies, such as those by Taylor et al,¹⁶ have tended to assume that tetrameric aggregates were a prevalent species within systems where clustering of alcohol molecules occurs. The tetrameric species is generally assumed to be prevalent because of its enhanced stability when it adopts a cyclic structure; this low-energy configuration has been verified in multiple studies of methanol through *ab initio* molecular orbital calculations,¹⁷ although not in mixtures or clusters including CO₂.

Zhang, Han & Xu¹⁸ carried out both *ab initio* quantum calculations and Monte Carlo simulations to sample the optimal geometries of a number of complexes with CO₂. These simulations as well found a weak stabilizing Lewis acid-Lewis base interaction between CO₂ and methanol, consistent with the attractive interactions suggested in the IR studies discussed above.

Monte Carlo Simulations

Monte Carlo simulation studies have probed the equilibrium structures of these mixtures. Stubbs and Siepmann¹⁹ determined that, paralleling experimentally determined trends, the fraction of ethanol monomers in supercritical CO₂ decreases with increasing temperature and exhibits very little dependence on pressure. Additionally, the sizes of hydrogen-bonded aggregates were calculated, and, although generally larger aggregates were less common than smaller aggregates, an unusual abundance of ethanol tetramers and pentamers was found. About 20-30% of these oligomers were calculated to be cyclic in structure.

Xu, Yang, and Hu²⁰, in a similar MC study of ethanol in CO₂ at a range of compositions, temperatures, and pressures, reported similar trends of more hydrogen

bonding with increasing temperature, less H-bonding with increasing pressure (although this change was minimal above 20MPa, echoing Stubbs and Siepmann's findings¹⁹) and more H-bonding with increasing ethanol mole fraction. They similarly found a prevalence of ethanol tetramers and pentamers, with the prevalence of cyclic structures increasing with increasing pressure and ethanol concentration. H-bond chain structures reached their maximum prevalence (at 328K and 20MPa) around 30% ethanol and cyclic structures predominated at higher ethanol mole fractions.

MC simulations carried out by Idrissi et al.²¹ focused on the thermodynamic properties of sc-CO₂/cosolvent mixtures as a means of validating various potential models, and they found that molar volume, molar free energy, and molar entropy in CO₂/ethanol mixtures at 313K reach a maximum around 0.05 mole fraction, which they attributed to the development of a liquid phase as ethanol molecules clustered together. (This temperature, where a two-phase system was observed to develop, is lower than those of all of the simulations discussed in this dissertation.)

Molecular Dynamics Simulations

Skarmoutsos, Dellis, and Samios²² conducted a molecular dynamics simulation of a 10% ethanol mole fraction mixture at temperature and pressure greater than the critical point. In addition to confirming the structural characteristics of hydrogen bonding observed in the Monte Carlo simulations described above, they were also able to probe the dynamics of the two components within the mixture. In particular, they observed that the motion of ethanol molecules was significantly slower compared to that of carbon dioxide molecules – ethanol molecules exhibited smaller diffusion

coefficients, longer reorientational relaxation times, and saw slower local environment reorganization.

The inhibited motion of ethanol was attributed to hydrogen-bonding phenomena; several observations in particular support this conclusion. First, reorientational relaxation times can be calculated along several vectors within the ethanol molecules, including each bond vector and the molecular dipole vector. Among these, rotation along the O-H bond vector was observed to be the slowest. Additionally, in a further study of the same system, Skarmoutsos et al.,²³ calculated separate orientational time correlation functions for ethanol molecules as a function of the number of hydrogen bonds each molecule participated in; ethanol molecules not participating in any hydrogen bonds (monomers) exhibited the fastest orientational relaxation, followed by ethanol molecules participating in one hydrogen bond, then those participating in two hydrogen bonds.

The above simulations from Samios' group have explored dynamics at only one thermodynamic state for the CO₂/EtOH system. An earlier study by Chatzis and Samios²⁴, however, probed the dynamics of CO₂/MeOH mixtures at a range of pressures. This study revealed similar evidence of hydrogen bonding to the studies above, with moderate changes in hydrogen bonding statistics with pressure, although the study did not note any consistent trends. Additionally, the translational dynamics of both components were examined, and it was noted that, although the diffusivities of both species decreased with increasing pressure, the mobility of CO₂ decreased more rapidly than that of MeOH.

Molecular dynamics studies by Aida, Aizawa, Kanakubo, and Nanjo²⁵ strictly examined structural properties of various mixtures of short-chain alcohols in

supercritical CO₂. For all short chain alcohols (methanol through butanol) pentamers were the most common cluster size; for ethanol, the fraction of clusters that were cyclic decreased with increasing ethanol mole fraction. Finally, self-diffusion coefficients of both components decreased with increasing alcohol mole fraction, and decreased modestly with increasing alcohol chain length.

Most recently, Reiser, McCann, Horsch, and Hasse²⁶ were able to use molecular dynamics simulation data – particularly, the fractions of ethanol molecules in three different hydrogen-bonding environments (monomers, donors, and acceptors) – to describe a large set of experimental ¹H NMR data from CO₂/ethanol mixtures at a wide range of temperatures, pressures, and concentrations, illustrating how relatively simple potential models can effectively represent the structure and properties of such mixtures.

Solutes in sc-CO₂/cosolvent systems

The above studies describe work done to understand CO₂/alcohol mixtures, yet, implicit in the description of short-chain alcohols as cosolvents is the understanding that these systems are of interest because of their interactions with solutes. Although little simulation work has been done in the supercritical regime, the area of this dissertation, there has been recent interest in connecting simulation and experimental work in the gas-expanded liquid regime. Compared with systems in the supercritical regime, gas-expanded liquid systems are typically at similarly high pressures, but at lower temperatures (<300K) and contain high mole fractions of the polar component.

Maroncelli's group has studied both the two solvent components independently of a solute, finding trends from simulation that echo results discussed above from the supercritical regime – slower diffusion of both components with increasing mole fraction of the polar component, for example.²⁷ When solutes are introduced into these

simulations, composition enhancement of the polar solute within with first solvation shell around the solute is also found, attributable to favorable hydrogen-bonding interactions; these trends are corroborated by solvatochromic experiments that show similar enrichment in the first solvation shell.²⁸ Although not involving an alcohol cosolvent, this work has extended to the description of the rearrangement of the first solvation shell (containing CO₂ + CH₃CN as the solute mixture) around a solute molecule that is excited as a solvatochromic probe. The dynamics of the first solvation shell, as measured through the time evolution of the solvatochromic shift upon excitation of the solute, also agree qualitatively with simulations, which show that excitation of the first solvation shell results in two types of rearrangement: first, a short timescale reorientation of solute molecules, followed by a slower reorganization of the solute as molecules move into and out of the first solvation shell.²⁹ Related simulation studies, involving a range of CO₂/cosolvent mixtures and solute molecules, have described site-specific behaviors such as preferential clustering of one solvent species around particular portions of a solute molecule,³⁰ and other experimental work on similar systems in the gas-expanded liquid and supercritical regimes has focused heavily on estimation of local solvent compositions based on solvatochromic shifts.^{6b, 9a, 31}

8.2 Central Questions and Dissertation Structure

While the above studies have established some interesting characteristics of these mixtures – the presence of an attractive CO₂/EtOH interaction, the prevalence of cyclic tetramers and pentamers, trends in dynamics and structure with thermodynamic conditions, and preferential solvation of solute molecules, these studies have not

explicitly linked the overall dynamical behavior to particular variations in structure that also occur as thermodynamic parameters are varied.

Our central questions, therefore, center around the following: In sc-CO₂/ethanol mixtures, how do changes in both structure and dynamics as conditions are varied combine to result in changes to the dynamical behavior of these systems? To this end, in the following chapters, we examine how both structure and dynamics of CO₂/ethanol mixtures, with and without solutes present, change under different thermodynamic conditions, varying temperature, pressure, and composition. In particular, we will examine the variations in local environments around solute and cosolvent molecules: How do the local environments around each solvent and solute molecule vary – both in their frequency within the mixture and over time under different thermodynamic conditions, and in their dynamical properties? To this end, we have carried out molecular dynamics simulations of sc-CO₂/ethanol mixtures, both with and without solutes, and analyzed their output. MD simulations are uniquely suited to help address questions about molecular-level dynamics, offering time-dependent data and direct access to molecular-level properties that can usually only be estimated or approximated from experimental measurements.

Chapter 2 examines these questions for sc-CO₂/ethanol mixtures without solutes, focusing particularly on the role of the hydrogen-bonded networks of ethanol molecules in governing the dynamics of the two cosolvent species. In Chapter 3, solutes are introduced into several sc-CO₂/ethanol mixtures, and we similarly study the dynamics of both the solute molecules and the structure and dynamics of their local environment, examining how solute characteristics affect solute/solvent interactions and the manner

in which they hinder or facilitate the movement of both. In Chapter 4, we summarize our findings and discuss further avenues for investigation in this area.

Chapter 9: Structure and Dynamics of Ethanol Aggregates in Supercritical CO₂/Ethanol Mixtures

In contrast with liquid organic solvents, supercritical carbon dioxide possesses many attractive properties, and is in widespread use in processes ranging from extraction of solids from foodstuffs to synthesis of nanostructured materials to analytical and industrial separations.³² The supercritical (sc) regime is attractive because it offers liquid-like densities but gas-like viscosities; additionally, near the critical temperature, thermodynamic properties of the mixture such as density, and properties of the solute such as mobility and solubility can be easily tuned with relatively small changes in temperature, composition, and pressure, affording the potential for control of chemical processes. Carbon dioxide is also readily available, easily recyclable, non-toxic, non-flammable, and relatively benign environmentally; moreover, its critical temperature – 304 K, or 31°C – is readily accessible under experimental conditions.

However, many polar organic compounds of interest for separation or extraction processes have poor solubility in non-dipolar sc-CO₂. Because of this difficulty with using an otherwise desirable supercritical solvent, polar cosolvents are often added to sc-CO₂ systems to enhance the solubility of polar compounds. Ethanol, which, like CO₂, is relatively benign and inexpensive, is among the more common cosolvents used in conjunction with CO₂,³³ and its addition (as with other small chain alcohols) has been seen to increase the solubility of organic components by several orders of magnitude^{5, 8} and to significantly increase solute diffusion coefficients.³⁴ Both solubility and solvent mobility have also been seen to vary appreciably as a function of the composition of the

solvent mixture, opening another avenue for tuning and control of chemical processes beyond variation of temperature and pressure,

As suggested by the use of the term “entrainer” to describe these cosolvents, it is believed that the enhanced solubility of polar compounds in scCO₂-cosolvent systems arises from a preferential interaction between the solute species and the polar cosolvent. Just as local heterogeneities and variations in local density exist within supercritical fluids,^{6a} the structure of the solvent/cosolvent mixture is also subject to local heterogeneities as each component associates preferentially with similar molecules. Spectroscopic studies have shown that a polar solute is preferentially surrounded by the polar cosolvent species.^{9b, 35} Similarly, individual ethanol molecules within these mixtures are likely to also be preferentially surrounded by other ethanol molecules.

To date, this understanding has been advanced through both experimental and computational studies, which have confirmed the expected mutual attraction between molecules of small chain alcohols through conventional hydrogen-bonding interactions both via spectroscopy^{12, 36} and via Monte Carlo and Molecular Dynamics (MD) simulation^{15, 19-20, 22, 25}. Some of these studies, both experimental¹³⁻¹⁴ and simulation^{20, 22-23} have also detected an attractive interaction between ethanol and carbon dioxide within these mixtures, resulting in local density enhancements of CO₂ in the vicinity of ethanol. In order to ultimately advance our understanding of how solvation, transport processes, and reactions take place within these supercritical solvent mixtures, and to optimize solvation parameters for various analytical and industrial processes, it will be necessary to not only to understand the interaction of each cosolvent with a potential solute, but the interactions of solvent molecules with one another.

In addition to differing intermolecular interactions, ethanol and carbon dioxide are also observed to have different dynamical properties within the same mixture. Previous MD studies of^{22-23, 25} CO₂/ethanol systems have reported that motion of ethanol molecules is significantly slower than that of carbon dioxide molecules – exhibiting smaller diffusion coefficients, longer orientational relaxation times, and saw slower local environment reorganization. This retardation of dynamics is attributed in part to the prevalence of alcohol-alcohol hydrogen bonding within these systems, but the direct connections between local hydrogen bonding and dynamics have not been directly examined.

As thermodynamic conditions are altered, both the structure and dynamics of the mixtures change. With respect to the predominant ethanol-ethanol interaction, both IR and NMR experiments^{12, 15} and molecular simulations^{20, 25} of CO₂/alcohol mixtures corroborate an increase in the fraction of molecules engaged in hydrogen bonding with increasing alcohol mole fraction. A subset of these studies^{15, 20} have also examined the effect of temperature and pressure, finding that the number of hydrogen bonds decreases with increasing temperature, as thermal energy disrupts hydrogen-bonding interactions, and increases modestly with increasing pressure.

The fact that ethanol molecules can participate in hydrogen bonds simultaneously as both donor and acceptor means that ethanol molecules can aggregate in clusters larger than dimers. It has generally been found that cyclic tetramers and pentamers are particularly common, with cyclic geometry of aggregates becoming increasingly prevalent as the concentration of the alcohol component increases^{19 20 25}. Although snapshots of molecular configurations offer a window into these hydrogen-bonding structures, dynamical simulations can elucidate the relationships between

these structures and the time-dependent behavior of ethanol molecules. The interactions between polar ethanol molecules may also be a model for the interactions between ethanol and polar or hydrogen-bonding solutes.

To date, although the dynamics of individual components within the mixture have been studied, and the overall structure of these mixtures has been characterized, there has been relatively little work seeking to connect the dynamics of each component with the local structure environment or to characterize how differences in structure relate to differences in dynamics at different thermodynamic conditions. This study therefore employs molecular dynamics simulations to explore the structure and dynamics of both cosolvents and of hydrogen-bonded ethanol aggregates across a range of densities and concentrations at a point just above the critical temperature, attempting to make connections between local structure and dynamical behavior, and to make contact with past experimental and theoretical studies where appropriate.

9.1 Simulation Parameters and Conditions.

Molecular dynamics simulations on ethanol/carbon dioxide mixtures at a range of compositions and pressures were conducted using DL_POLY 2.0,³⁷ using the Verlet leapfrog algorithm to solve the equations of motion. All simulations were conducted using cubic periodic boundary conditions, using the standard Ewald summation with conducting boundary conditions³⁸ to treat long-range electrostatic interactions; an Ewald truncation radius of 12 Å was used in all simulations. Simulations at seven different thermodynamic states were conducted, forming three series in which either temperature, pressure, or composition were varied while the other thermodynamic parameters were held constant: (a) a composition series conducted at 338 K and about

20 MPa, containing either 5, 10, or 15 mole percent ethanol; (b) a temperature series conducted at about 20MPa and 10 mole percent ethanol, at either 328K, 338K, or 358K; and (c) a pressure series conducted at 338K and 10 mole percent ethanol, at either 20 MPa, 30 MPa, or 40 MPa. Each simulation was conducted with a total of 500 molecules, either 25, 50, or 75 of which were ethanol molecules, depending on the relevant composition. Although the critical temperature of CO₂ lies at 304 K, the addition of ethanol at mole fractions of between 5 and 15% results in an elevation of the critical point; for these mixtures, the critical temperature has been estimated to lie between 310 K and 325K, and the critical pressure is estimated to be between 9 and 13 MPa, placing the simulation conditions within the supercritical regime.³⁹

Carbon dioxide was simulated with the EPM2 (Elementary Physical Model) force field. This potential represents each site with superimposed Lennard-Jones and Coulombic potentials, and imposes fixed bond lengths and a rigid linear geometry. The EPM2 model was chosen for its ability to accurately simulate carbon dioxide's critical properties and liquid-vapor coexistence curve; it reproduces CO₂'s critical temperature and pressure within simulation uncertainty and the critical density within 3% of experimental values.⁴⁰ Ethanol was modeled using the OPLS-UA (Optimized Potentials for Liquid Systems – United Atom) potential parameters developed by Jorgensen.⁴¹ This set of force fields, chosen for its ability to reproduce hydrogen bonding behavior between alcohol molecules,³⁶ represents both the methyl and methylene groups of ethanol as united Lennard-Jones + Coulomb sites. Oxygen is also represented with a Lennard-Jones potential and point charge, while the terminal alcohol hydrogen is represented simply as a point charge. This force field includes fixed bond lengths and

bond angles, and the SHAKE algorithm⁴² is used to handle bond constraints. Rotation about the dihedral angle φ is governed by the torsional potential:

$$U(\varphi) = U_0 + \frac{1}{2}U_1(1 + \cos \varphi) + \frac{1}{2}U_2(1 + \cos 2\varphi) + \frac{1}{2}U_3(1 + \cos 3\varphi) \quad (1)$$

Each simulation was conducted with a 0.5 fs timestep. The Nosé-Hoover thermostat⁴³ was employed to control temperature; and the Nosé-Hoover barostat⁴⁴ was used to control pressure during the first phase of simulations, both with frequencies of 0.5 ps⁻¹.

Equilibration Overview

The simulation trajectories were obtained using a method similar to that described by Roney et al.⁴⁵ – beginning with a simulation box containing the desired composition, the system was first equilibrated with respect to pressure in the NPT ensemble, allowing the simulation box size to adjust to yield the appropriate density. Using this data, a starting configuration of an appropriate volume was selected to conduct a simulation in the NVT ensemble at the composition and pressure of interest.

In the first phase of the simulation, an initial set of atomic coordinates was generated in which ethanol and carbon dioxide molecules were randomly distributed across a face-centered cubic lattice, an initial configuration chosen so as to minimize the possibility of overlap between adjacent molecules. This system was allowed to equilibrate for 200 ps in the NPT ensemble, using the Nosé-Hoover thermostat and barostat to regulate temperature and pressure. A second run of 300 ps was then conducted to determine an average simulation box size for the system. A starting configuration sampled from this second run, chosen to be within 1% of the calculated average volume, was selected as the starting point for a production run in the NVT

(canonical) ensemble. These simulations were first allowed to equilibrate for 100 ps, and then a production run of 400 ps was conducted. From these simulations, pressures were also calculated to verify that the NVT simulation reasonably approximated the desired conditions. These pressures, along with the simulation cell density of each system, are reported in Table 9.1. In each case, the calculated pressure was determined to be between 83% and 97% of the desired pressure.

Table 9.1. Simulation parameters for NVT simulation runs.

Ethanol Mole Percentage	Desired Pressure (MPa)	Temperature (K)	Density (g/cm³)	Calculated Pressure (MPa)
5.00%	20	338	0.693	18.4
10.00%	20	338	0.701	17.2
10.00%	30	338	0.792	27.0
10.00%	40	338	0.847	37.5
15.00%	20	338	0.712	16.7
10.00%	20	328	0.740	16.1
10.00%	20	358	0.656	18.4

9.2 Structure of CO₂/Ethanol Mixtures

9.2.1. Distribution Functions

Intermolecular structure for these mixtures was analyzed using a variety of methods. Initially, site-site radial distribution functions and local mole fractions were calculated based on configurations of atomic coordinates generated during each simulation run.

Local mole fractions, $\zeta_{BA}(r)$ for both species, around molecules of a given type, are calculated as the fraction of molecules of that type (out of all molecules) within a given distance from the molecule's center of mass.⁴⁶ This can be expressed as:

$$\zeta_{BA}(r) = \frac{n_{BA}(r)}{n_{AA}(r) + n_{BA}(r)} \quad (1)$$

where $n_{BA}(r)$ is the average local coordination number for the center of mass of molecules of type B around the center of mass of molecules of type A, calculated as:

$$n_{BA}(r) = \int_0^r r^2 g_{BA}(r) \rho dr \quad (2)$$

based on the pair correlation function, $g_{BA}(r)$ and bulk molecular density, ρ .

Characteristic peaks of H-bonding.

Figure 1a illustrates several of the ethanol site-site RDF's, which exhibit a characteristic signatures of hydrogen bonding: pronounced peaks within the O-H RDF at 1.9 Å, the H-H RDF at 2.3 Å, and the O-O RDF at 2.8 Å. These peak positions, which are consistent with a roughly linear O–H···O geometry, remain stable between simulations despite changes in composition and pressure. The second peak discernable in the O-H RDF reflects the presence of additional ethanol molecules in the second coordination shell of the central ethanol molecule, hydrogen-bonded to nearest neighbors in the first shell; this is an indicator of the existence of ethanol aggregates of more than two hydrogen-bonded molecules.

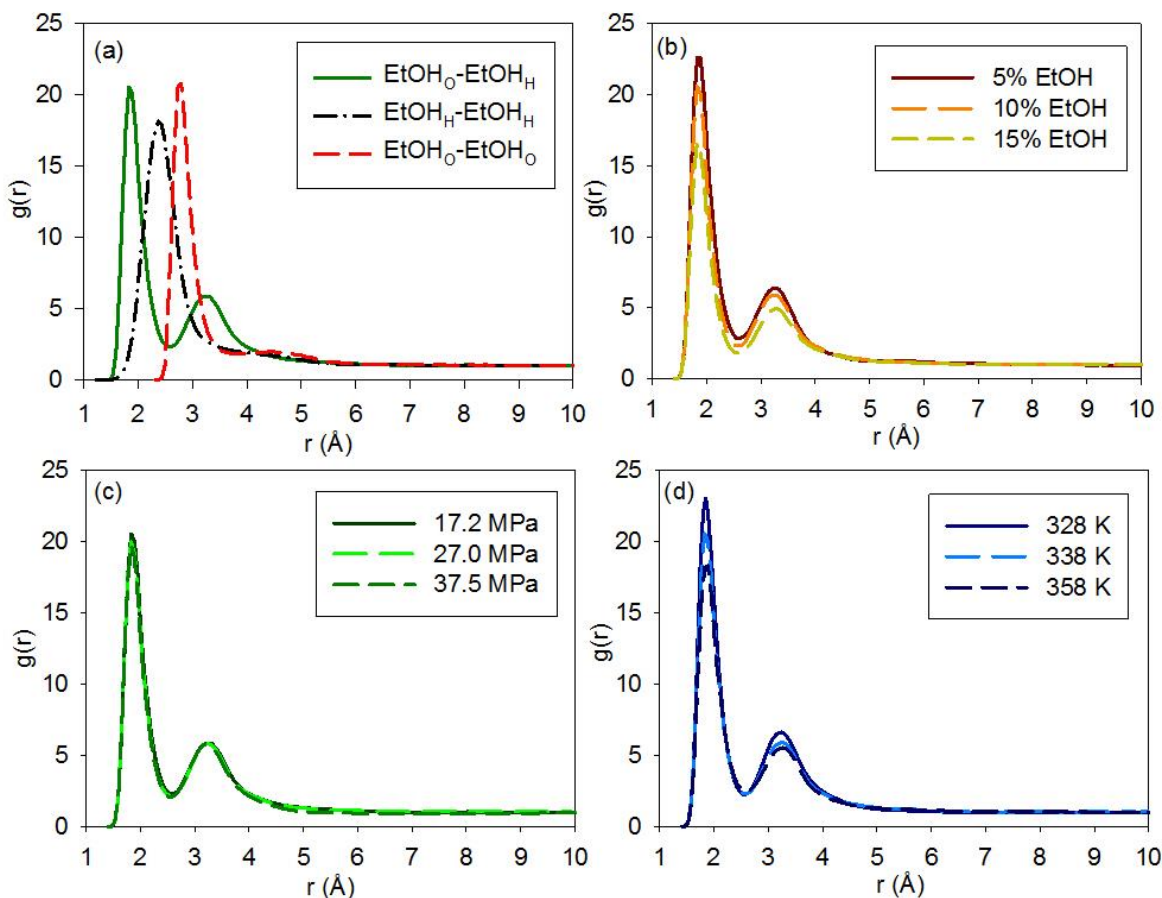


Figure 9.1. (a) Ethanol site-site RDF's for hydroxyl group O-H, H-H, and O-O interactions within the 328K, 10% EtOH and 17.2 MPa simulation; (b) O-H site-site RDFs for simulations at 17-18 MPa and 5%, 10%, and 15% EtOH mole fraction; (c) O-H site-site RDFs for simulations at 10% EtOH mole fraction and 17.2 MPa, 27.0 MPa, and 27.5 MPa. (d) O-H site-site RDFs for simulations at 328K, 338K, and 358K, with 10% ethanol and 16-18 MPa.

Effects of composition.

It should be noted that the comparatively high first maxima of these RDF's (for the O-H peak, 22.6 for 5% ethanol mole fraction, 18.4 MPa, and 338 K) arise as a result of the normalization of the local ethanol density with respect to the long-distance ethanol density, so the changes in the maximum of the first peak (from 22.6 at 5%

ethanol mole fraction to 16.6 at 15% ethanol) with composition differences do not directly reflect the differences in local composition between these two mixtures. Local mole fraction plots (Figure 9.2) are more informative with respect to the local composition enhancement; within the simulation at 5% ethanol, the maximum local composition enhancement (at about 5 Å) is approximately three times the bulk mole fraction, whereas the maximum local mole fraction within the 15% ethanol mixture is only about twice the bulk value. Although the number of ethanol molecules participating in the first solvation shell increases, the composition enhancement relative to the bulk concentration decreases with increasing ethanol mole fraction.

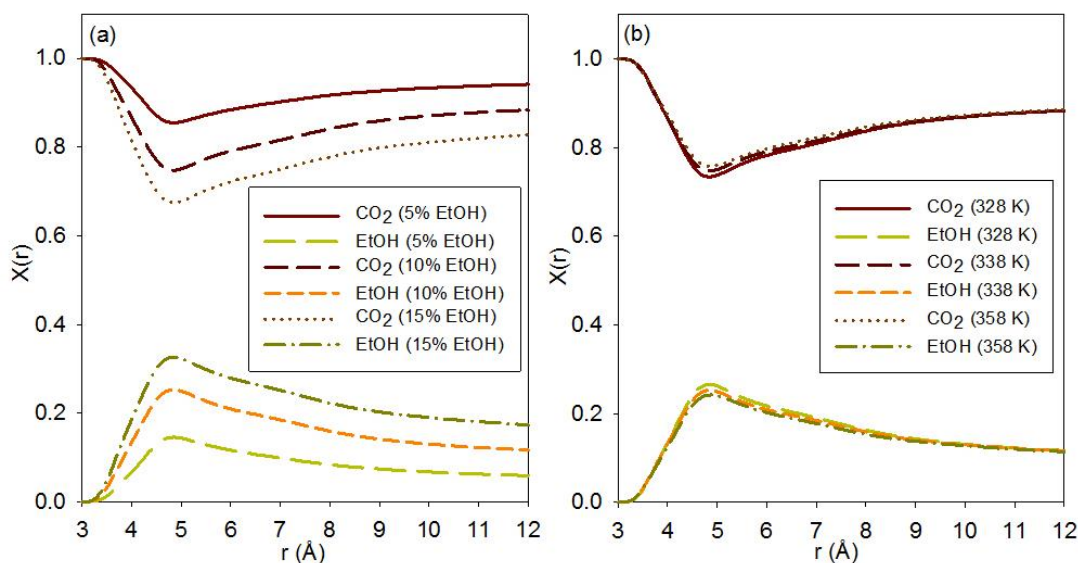


Figure 9.2. Center-of-mass local mole fraction of CO₂ and ethanol around ethanol as a function of (a) composition and (b) temperature.

Effects of pressure and temperature.

As pressure increases from 17.2 MPa to 37.5 MPa at a constant composition of 10% ethanol, relatively small changes are seen in ethanol-ethanol association. The site-

site RDFs characteristic of the hydrogen-bonding interaction decrease minimally with increasing pressure, (Figure 9.1c), the local composition enhancement remains essentially constant (around 0.32 maximum local mole fraction), reflecting the high stability of EtOH-EtOH interactions in the face of increasing pressure for carbon dioxide molecules to penetrate into the first solvation shell. Xu et al.²⁰ reported similarly small effects of pressure changes between 20 and 40 MPa on ethanol-ethanol RDFs, and comparable results have been seen for the same pressure range in simulations of methanol-CO₂ mixtures.²⁴ With increasing temperature, the maximum within O-H RDFs decreases slightly in intensity, as does the maximum local composition enhancement (decreasing from 0.266 ethanol mole fraction at 328 K to 0.242 ethanol mole fraction at 358 K), reflecting an expected decrease in the degree of ethanol-ethanol association with increasing temperature.

EtOH-CO₂ interactions.

Both Skarmoutsos²² and Xu et al.²⁰ described a hydrogen-bonding-like interaction between ethanol and CO₂, evidenced by features such as the shoulder at 2.4 Å present in the RDFs between the H_{EtOH}-O_{CO₂} RDF (Figure 9.3a). This interaction appears to decrease in prevalence with increasing ethanol mole fraction (Figure 9.3a), as more ethanol molecules are available for the more highly favored ethanol-ethanol interactions. Conversely, Figure 9.3b shows that as ethanol-ethanol interactions are disrupted at higher temperatures, the shoulder indicative of this ethanol-CO₂ arrangement grows in intensity. However, changes in ethanol-CO₂ interactions with pressure are minimal, paralleling the similar lack of variation in the pressure series of ethanol-ethanol RDFs.

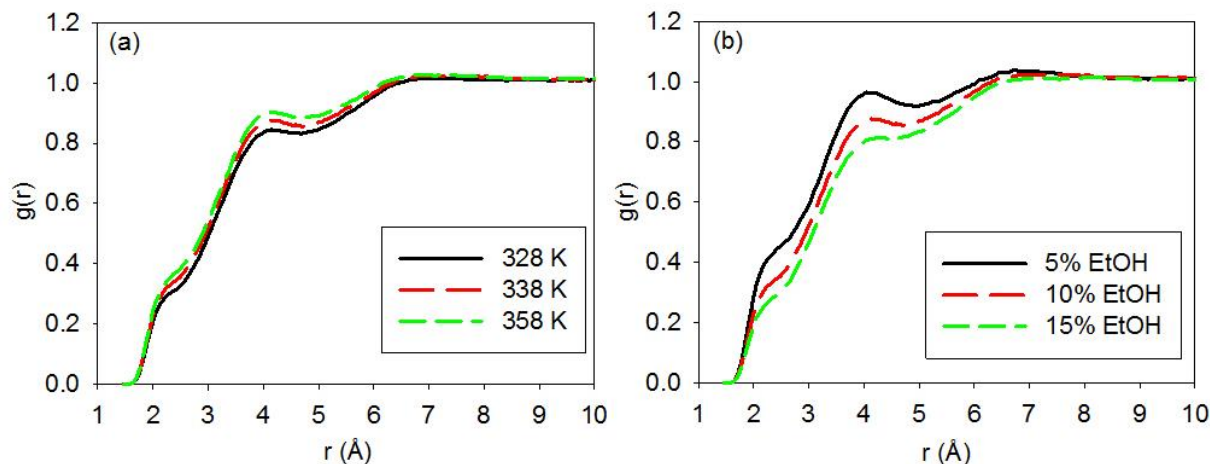


Figure 9.3. $H_{\text{EtOH-O}_{\text{CO}_2}}$ RDFs as a function of (a) mixture temperature and (b) composition.

9.2.2 Hydrogen Bonding Patterns

The presence of clear indicators of significant hydrogen bonding in RDFs suggests a need to analyze hydrogen bonding interactions within these systems specifically. Differing criteria have been proposed to define whether hydrogen bonds exist between neighboring molecules, and among the most broadly used is a geometric criterion.⁴⁷ Under criteria frequently used for simulations of small-chain alcohols in CO_2 ,^{15, 20, 23-24} two neighboring molecules are considered to be engaged in hydrogen bonding if three geometric criteria are met with respect to the O-H---O intermolecular geometry: an oxygen-oxygen-hydrogen angle of less than 30° , an intermolecular oxygen-hydrogen distance of less than 2.6 \AA , and an intermolecular oxygen-oxygen distance of less than 3.5 \AA . These particular distances have been observed to be in good agreement with energetic calculations,³⁶ to be valid at a range of conditions within the supercritical region,¹⁸ and are also consistent with the first minima of the radial distribution functions in Figure 9.1.

Number of hydrogen bonds per ethanol molecule.

Because individual ethanol molecules can participate in hydrogen bonding both as an electron donor and acceptor, an ethanol molecule can act as an acceptor for multiple hydrogen bonds – in this and other studies, as many as several percent of ethanol molecules are calculated to participate in three hydrogen bonds,^{20, 23} and in this study, a very small number of ethanol molecules were calculated to participate in 4 hydrogen bonds. The fraction of ethanol molecules engaged in n hydrogen bonds and the average number of hydrogen bonds per molecule are presented in Table 9.3. The number of hydrogen bonds per molecule increases both with increasing ethanol mole fraction and increasing pressure, and decreases with increasing temperature, consistent with the patterns described above from local density and radial distribution plots. Similarly, the number of ethanol molecules existing as “monomers” – not hydrogen-bonded to any other ethanol molecules – decreases with increasing ethanol concentration and pressure. Studies conducted at lower temperatures²⁰ have observed similar patterns of hydrogen bonding with composition and pressure, albeit with a greater amount of hydrogen bonding than we describe here, as those simulations were conducted with lower temperatures.

Hydrogen-bonded aggregates.

Networks of hydrogen-bonded ethanol molecules can be considered to form aggregates, whose size and structure can be described by considering ethanol molecules and hydrogen bonds as vertices and edges of a graph, respectively. This graph can be used to analyze structure of hydrogen-bonded aggregates, particularly with respect to

the size of networks of linked ethanol molecules and their connectivity – whether they possess linear, cyclic, or other structures. The fraction of ethanol molecules participating in aggregates of sizes $n=1$ (ethanol monomers) through $n=8$ is shown in Figure 9.4 for the series of simulations at varying composition, pressure, and temperature; average aggregate size for all simulations is reported in Table 9.2. Trends in aggregate size distributions parallel trends in hydrogen bonding activity, with average aggregate size increasing as pressure increases, as temperature decreases, or as ethanol mole fraction increases.

Although the fraction of molecules participating in a cluster of size n would be expected to decrease with increasing n , there is one notable exception to this trend: the unusual prevalence of ethanol tetramers and pentamers. This has previously been described by MD and MC studies of similar mixtures; at lower temperatures or higher ethanol concentration than those described in this study, the number of ethanol molecules participating in tetramers or pentamers can exceed the number participating in trimers or even dimers^{20, 25}. This suggests that, perhaps because of their structure, clusters of this size may be particularly stable and long-lived. In fact, *ab initio* quantum calculations have predicted that cyclic tetramers of ethanol should be particularly stable,⁴⁸ and the effect of cyclic structures on dynamics is discussed in section 9.3.3 below.

Table 9.2. Structural characteristics related to hydrogen bonding for all series of simulated mixtures.

Temperature (K)	EtOH mole fraction	Pressure (MPa)	Fraction in 0 hb	Fraction in 1 hb	Fraction in 2 hb	Fraction in 3 hb	Average number of h-bonds	Average cluster size	Fraction of cyclic clusters of 4 or more molecules
338	5.0%	18.4	0.594	0.279	0.119	0.007	0.538	1.88	28.9%
338	10.0%	17.2	0.462	0.338	0.188	0.011	0.747	2.32	19.8%
338	15.0%	16.7	0.370	0.370	0.244	0.016	0.910	2.82	18.5%
338	10.0%	17.2	0.462	0.338	0.188	0.011	0.747	2.32	19.8%
338	10.0%	27.0	0.478	0.322	0.188	0.011	0.735	2.31	30.0%
338	10.0%	37.5	0.386	0.339	0.165	0.010	0.699	2.16	36.1%
328	10.0%	16.1	0.415	0.361	0.212	0.012	0.821	2.44	27.4%
338	10.0%	17.2	0.462	0.338	0.188	0.011	0.747	2.32	19.8%
358	10.0%	18.4	0.399	0.338	0.154	0.009	0.673	2.06	29.4%

With this result in mind, cluster geometries were analyzed for the presence of cyclic structures, and the fraction of aggregates of tetramer size or larger that contain cyclic structures is also reported in Table 9.2. Generally, the fraction of aggregates containing cyclic structures decreases with increasing ethanol mole fraction; this contrasts with previous studies at lower temperatures in which the number of aggregates containing cyclic structures was seen to increase with increasing mole fraction⁴⁹, or not to vary significantly with changes in composition.²⁵ With increasing

pressure, however, the fraction of ethanol aggregates containing cyclic structures increases markedly.

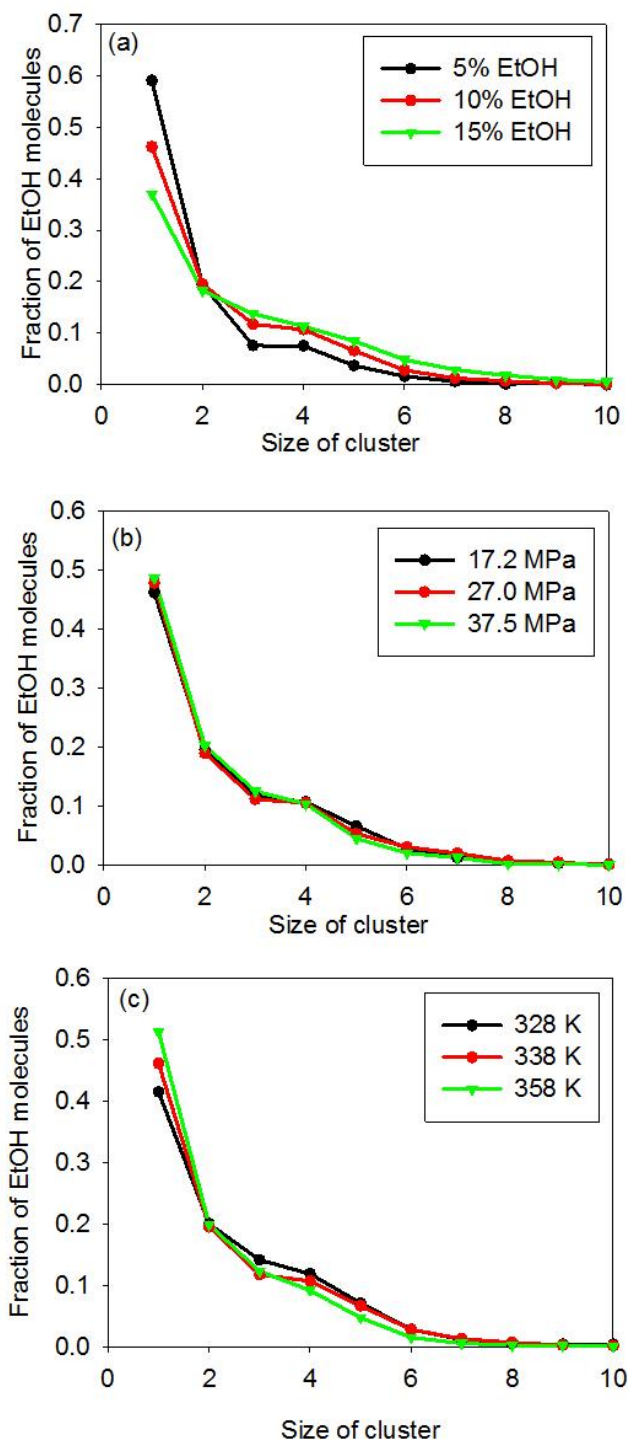


Figure 9.4. Fraction of ethanol molecules participating in aggregates of a given size within the mixture as a function of (a) ethanol mole fraction, (b) pressure, and (c) temperature.

9.3 Dynamical Properties

9.3.1 Diffusion

Molecular dynamics simulations afford the opportunity to study the translational motion of both species within CO₂/ethanol mixtures, how they differ from each other, and how both change as a function of thermodynamic conditions. Within the supercritical regime, diffusion coefficients vary significantly as a function of these parameters, and the presence of two different components within these mixtures has been observed to result in different behaviors for the two components, with the diffusion coefficient for a self-associating polar cosolvent calculated as being slower than that of non-polar, less strongly-associating CO₂ molecule of similar size within the same mixture. Below, we describe ways in which this phenomenon manifests as thermodynamic conditions are varied.

Self-diffusion coefficients can be calculated from mean squared displacements for each component in each mixture. The limiting slope of mean squared displacement with respect to time can be related to the diffusion coefficient via the Einstein relation:

$$\frac{\partial \langle |\Delta \mathbf{r}(t)|^2 \rangle}{\partial t} = 6D \quad (3)$$

Where $\Delta \mathbf{r}(t) = \mathbf{r}(t) - \mathbf{r}(0)$, with $\mathbf{r}(t)$ being the position vector for the molecular center of mass. The diffusion coefficients for both components in each simulation are provided in Table 9.3. Despite having a similar molecular weight and size as carbon dioxide, in every simulation, ethanol has a diffusion coefficient (D_{EtOH}) that is only 53% to 63% that of diffusion coefficient for CO₂ (D_{CO_2}) in the same mixture, reflecting a hindrance in translational dynamics brought about by the self-association of ethanol molecules.

Table 9.3. Dynamical characteristics of mixtures within each series of simulations.

<i>T</i> (K)	EtOH mole fraction	Pressure (MPa)	CO₂ diffusion coefficient (10⁻⁹ m²/s)	EtOH diffusion coefficient (10⁻⁹ m²/s)	CO₂ orientational relaxation time (ps)	EtOH orientational relaxation time (dipole vector, ps)	Intermittent hydrogen bond lifetime (ps)	Continuous hydrogen bond lifetime (ps)
338	5.00%	18.4	25.7	16.2	0.179	1.25	17.4	2.11
338	10.00%	17.2	24.5	14.9	0.188	1.80	18.1	2.07
338	15.00%	16.7	23.1	12.2	0.197	2.13	16.6	2.09
338	10.00%	17.2	24.5	14.9	0.188	1.80	18.1	2.07
338	10.00%	27.0	20.1	12.1	0.231	1.90	16.5	2.06
338	10.00%	37.5	16.6	10.5	0.264	1.83	15.2	2.03
328	10.00%	16.1	22.0	13.4	0.212	2.05	18.9	2.24
338	10.00%	17.2	24.5	14.9	0.188	1.80	18.1	2.07
358	10.00%	18.4	28.3	17.2	0.161	1.34	14.0	1.96

As a function of temperature and pressure, diffusion coefficients vary as would be expected within a single-component system – both components diffuse faster with increasing temperature and more slowly with increasing pressure. Moreover, changes in temperature and pressure, while they may have substantial effects on dynamics within the system, have only modest effects upon the relative magnitudes of the diffusion coefficients of the two species within the mixture. As temperature is varied, the $D_{\text{EtOH}}/D_{\text{CO}_2}$ ratio stays relatively constant, with the diffusion coefficient for ethanol in all three simulations of the temperature simulation at about 61% that for CO₂; as pressure

is varied, the ratio of diffusion coefficients vary slightly more, but is still between 60% to 63% that of CO₂. This is particularly remarkable in light of the substantial net changes in diffusion coefficients for both components, which vary by more than 25% across each series.

However, for the set of simulations in which composition varies, the dynamics of the two components do not follow patterns that are as closely parallel to each other. While both components diffuse more slowly with increasing ethanol concentration, ethanol diffusion coefficients change at a much faster rate than diffusion of CO₂. While $D_{\text{EtOH}}/D_{\text{CO}_2}$ ranges between 0.60 and 0.63 for the simulations conducted at 10% ethanol mole fraction, this ratio falls as low as 0.53 in the 15% ethanol conditions. Moreover, as composition is increased from 5% to 15%, D_{CO_2} decreases by 10.4% (from $25.75 \times 10^{-9} \text{ m}^2/\text{s}$ to $23.08 \times 10^{-9} \text{ m}^2/\text{s}$), while D_{EtOH} decreases by 24.3% (from $16.15 \times 10^{-9} \text{ m}^2/\text{s}$ to $12.23 \times 10^{-9} \text{ m}^2/\text{s}$). Changes in ethanol mole fraction have significant effects on both components, but have a greater effect on diffusion of ethanol molecules. These differences in D_{EtOH} as a function of composition are of comparable magnitude to the experimentally measured differences in diffusivity of solutes dissolved in mixtures of CO₂ and ethanol at similar conditions. For a composition contrast of similar scale (from 4.6% ethanol to 23.8% ethanol, at 20 MPa and 317 K), for instance, the diffusivity of dibenzyl ether in a sc-CO₂/ethanol mixture was observed to decrease 31%, from $6.90 \times 10^{-9} \text{ m}^2/\text{s}$ to $4.70 \times 10^{-9} \text{ m}^2/\text{s}$;³⁴ this is comparable to the 24.3% decrease in D_{EtOH} within a mixture as its mole fraction is increased from 5% to 15% in our simulations.

The diffusion of both molecules of polar solvents and of polar ethanol molecules themselves are believed to be heavily influenced by interactions with nearby ethanol molecules. As noted in the previous section, changes in composition yielded the most

substantial changes in mixture structure – in particular, changes in hydrogen bonding structure – and these changes in local environment are likely to affect an individual molecule’s motion.

Therefore, we hypothesize that the local hydrogen-bonding environment around each individual ethanol molecule is likely to significantly affect its diffusion through the mixture. To investigate this, we examined the diffusion of ethanol molecules as a function of the number of hydrogen bonds they participated in. For each sub-population of ethanol molecules – those participating in either 0, 1, or 2 hydrogen bonds – mean-squared displacements were calculated. (Note that unlike mean-squared displacements for the entire population of ethanol molecules, each subset varied in membership with time; displacement was only included in the mean square displacement calculation if the molecule participated in the same number of hydrogen bonds in every timestep the entire period from time 0 to time t . Because of the relatively short hydrogen-bonding lifetimes, reliable MSDs could only be calculated out to around 1 ps for these sub-populations.) Diffusion coefficients were then calculated from the slope of the mean squared displacement with respect to time.

Table 9.4 presents diffusion coefficients for ethanol molecules falling into each of these three categories within the simulation at 10% ethanol and 17.2 MPa, as well as the diffusion coefficients averaged over all molecules; Figure 9.5 presents the MSD plots from which these values are calculated. For inclusion in these mean-squared displacement calculations, molecules must maintain a consistent hydrogen bonding state between times 0 and t . These data reflect a consistent pattern of molecules engaged in a greater number of hydrogen bonds diffusing more slowly. In particular, the diffusion coefficient for non-bonded ethanol molecules is actually more than double that

of the sample of all ethanol molecules. Given that, according to the hydrogen-bonding statistics reported above, more than half of ethanol molecules participate in a hydrogen bond at any given time, this is consistent with the bulk diffusion coefficients for ethanol. Increasing numbers of hydrogen bonds also result in somewhat slower diffusion of ethanol within the system described in Table 9.4; the diffusion coefficient for ethanol molecules participating in 2 hydrogen bonds (and therefore necessarily part of hydrogen-bonded aggregates of three or more molecules) is only 65% of the rate of those which are part of only one hydrogen bond.

Table 9.4. Ethanol diffusion coefficients for various populations within the simulation at 10% ethanol mole fraction, 338 K, and 17.2 MPa.

EtOH population	EtOH diffusion coefficient (10^{-9} m²/s)
bulk	14.9
0 hydrogen bonds (cluster of size 1)	36.6
1 hydrogen bond	1.00
2 hydrogen bonds	0.64
cluster of size 2	1.56
cluster of size 3	1.23
cluster of size 4	1.18
cluster of size 5	0.61

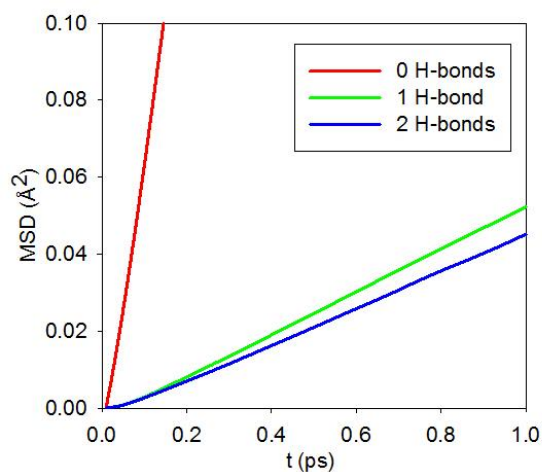


Figure 9.5. Mean-squared displacements as a function of time for ethanol molecules participating in 0, 1, or 2 hydrogen bonds, within the simulation at 10% ethanol mole fraction, 338 K, and 17.2 MPa.

Cluster-size dependent diffusion coefficients are also of interest, and are also reported in Table 9.4. These diffusion coefficients are also calculated using the Einstein relation from mean squared displacements (calculated out to a time of 10 ps); however, molecules are included in MSD calculations at time t only if they have remained in a particular state during every timestep between time 0 and time t – that is, they have continuously had a given number of hydrogen bonds or been a part of a cluster of a given size. As the size of hydrogen-bonded clusters increases, the corresponding diffusion coefficient for individual ethanol molecules within those clusters also decreases. As the hydrodynamic volume of a cluster increases with the addition of additional ethanol molecules, the diffusion of that cluster will slow correspondingly. If diffusion coefficients as a function of hydrogen bonding state are compared with values calculated as a function of cluster size, it also becomes apparent that diffusion is governed, in part, by cluster size. For instance, while an ethanol molecule participating

in at least two hydrogen bonds must be a part of a cluster of at least three molecules, the average diffusion coefficient for such molecules is actually most similar to that of ethanol molecules within clusters of 5 molecules, reflecting the frequency with which they participate in these larger clusters.

9.3.2 Orientational Time Correlations

Orientational time correlation functions, as measures of the mobility of the various components of these mixtures, show trends similar to those seen among diffusion coefficients. Previous research including MD studies by Skarmoutsos et al.²² at a single thermodynamic state have shown that carbon dioxide molecules rotate more freely within the mixture as compared with ethanol molecules, which have longer orientational relaxation times. This is also found within our simulations at all temperatures, pressures, and compositions. For ethanol and carbon dioxide molecules in our simulations, we have calculated first-order Legendre polynomial orientational time correlation functions:

$$C_{1,x}(t) = \frac{1}{N} \sum_{i=1}^N \langle \cos \theta_{i,x}(t) \rangle \quad (4)$$

where $\theta_{i,x}$ is the angle between vector \mathbf{x} embedded in the i^{th} molecule at time 0 and time t . Orientational relaxation times, calculated by integration of $C_{1,x}(t)$, with times beyond those calculated in the simulation handled by an exponential fit, are reported in Table 9.3 for the CO₂ bond vector and for the ethanol dipole vector. Figure 9.6 illustrates changes in orientational time correlation functions (TCFs) for ethanol with changes in both temperature and composition; changes in orientational TCFs with pressure were minimal by comparison.

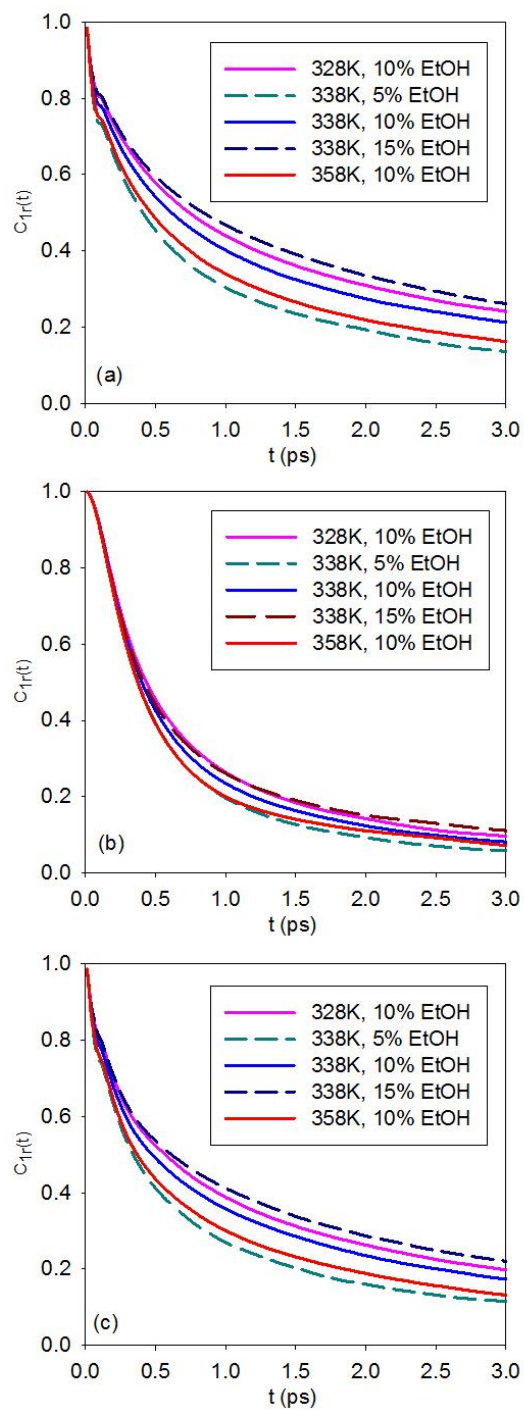


Figure 9.6. Orientational time correlation functions for ethanol molecules as a function of composition and temperature, for (a) the O-H bond vector, (b) the C-C bond vector, (c) the molecular dipole vector, within the temperature and composition simulation series (all at pressures between 16.1-18.4 MPa).

Effects of pressure.

Among the three thermodynamic variables explored, many of the effects on orientational relaxation are generally similar to those as on diffusion – faster orientational relaxation, evidenced by shorter characteristic reorientation times, is seen for both components with increasing temperature and decreasing ethanol mole fraction. The exception to these patterns is seen for reorientation of ethanol as a function of pressure; while CO₂'s orientational relaxation times slow from 0.188 ps to 0.264 ps as pressure increases from 17.2 MPa to 37.5 MPa (as shown in Figure 9.7), ethanol's relaxation times remain relatively constant, around 1.8-1.9 ps, and show no clear trend with respect to pressure. This, however, is consistent with the structural analyses reported earlier; along the composition series, ethanol-ethanol RDFs and hydrogen bonding statistics change relatively little, especially as compared with the changes apparent in the composition and pressure series.

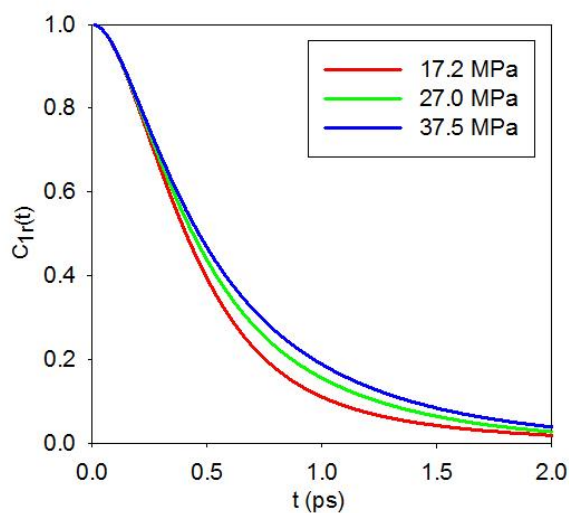


Figure 9.7. Orientational time correlation functions for CO₂ in the three simulations within the pressure series (all at 338 K and 10% ethanol).

Effects of composition.

Within the composition series, both CO₂ and ethanol's orientational relaxation is slowed as the ethanol mole fraction increases; however, compared to the differences in ethanol's dynamics, changes in the rate of reorientation time of carbon dioxide as a function of mixture composition are relatively small. Between the 5% ethanol mixture and the 15% ethanol mixture, orientational relaxation times lengthen from 0.179 ps to 0.197 ps for CO₂ – a change of about 9.1%, while ethanol orientational relaxation times lengthen from 1.25 ps to 2.13 ps – a change of 70.4%. Since CO₂ does not participate in intermolecular interactions with the strength of ethanol-ethanol hydrogen bonds, this modest slowing of rotation is likely linked to the presence of the favorable hydrogen-bond-like CO₂-ethanol interactions discussed above, of which there are a greater number in the presence of more ethanol molecules.

Effects of temperature.

Similarly, as temperature is increased from 328 K to 358 K, CO₂'s orientational relaxation times decrease by 24.0% while the corresponding parameter for ethanol decreases by 34.6%. While the increased thermal energy is no doubt responsible for a significant amount of the increase in rotational motion for both components, for ethanol, an additional hindrance has been removed in that fewer ethanol molecules participate in hydrogen bonds at higher temperature, and are therefore less constrained from rotation by neighboring ethanol molecules.

The orientational relaxation times discussed above apply to the ethanol dipole vector. However, in addition to this vector, reorientation times can be calculated for the

three bond vectors within the ethanol molecule, and these are reported in Table 9.5, for the population of ethanol molecules within the simulation at 10% ethanol mole fraction, 338 K, and 17.2 MPa. The slowest reorientation occurs along the O-H bond vector, which is the most affected by hydrogen bonding with nearby ethanol molecules, while the C-C bond vector, furthest from the sites participating in H-bonding, exhibits the fastest orientational relaxation. Because of the torsional degree of freedom within the ethanol molecule, reorientation of the C-C bond is clearly possible even when the C-C and O-H bonds are immobilized due to hydrogen-bonding.

Table 9.5. Orientational relaxation times for various vectors within subpopulations of ethanol molecules, for the simulation at 10% ethanol mole fraction, 338 K, and 17.2 MPa.

Bond	Orientational relaxation times (ps)			
	All molecules	Molecules participating in 0 h-bonds	Molecules participating in 1 h-bond	Molecules participating in 2 h-bonds
O-H	2.19	0.33	2.08	8.83
C-O	2.16	0.38	1.36	5.06
C-C	1.12	0.27	0.60	1.08

Evidence of the effects of hydrogen bonding is also clearly seen when rotational dynamics are looked at as a function of local environment. Additionally, Table 9.5 reports the reorientation times for subpopulations of ethanol molecules within this simulation – those participating in 0, 1, or 2 hydrogen bonds. As with diffusion, molecules are only included within each population if they continuously maintain their

hydrogen-bonding state throughout the period in question; this results in orientational relaxation times reported as a function of hydrogen bonding state here being substantially longer than those reported by Skarmoutsos et al.,²³ who divided their populations simply by initial hydrogen bonding state and did not remove molecules from the population as they broke or formed additional bonds. As can be seen, ethanol molecules which participate in fewer hydrogen bonds show faster orientational relaxation along all three bond vectors, although the change is greatest with respect to the O-H bond vector (0.330 ps for ethanol molecules participating in zero hydrogen bonds, 8.828 ps for ethanol molecules participating in two hydrogen bonds. For molecules engaged in no hydrogen bonds, the reorientation times along all three bond vectors are comparable, although reorientation along the central C-O bond vector is slowest.

These orientational relaxation times can be used to examine the origin of the differences in dynamics between mixtures with different compositions. Figure 9.8 plots orientational TCFs along the O-H bond vector for ethanol molecules participating in 0, 1, and 2 hydrogen bonds, respectively, within the three mixtures with different compositions at 338 K and similar pressures (17-18 MPa). As would be expected, ethanol molecules engaged in greater numbers of hydrogen bonds show substantially slower orientational relaxation. However, reorientational dynamics of ethanol molecules in similar hydrogen bonding environments are strikingly similar despite the differences in composition. This suggests that the differences in overall rotational dynamics between mixtures in the composition series arise due to the differing distributions of local environments, rather than due to some other longer-scale interactions. Additionally, these plots make clear a short-time oscillatory feature associated with

librational motion of ethanol molecules within each hydrogen-bonded state, a feature, commonly seen in such autocorrelation functions for liquid ethanol.⁵⁰ The maximum associated with this librational motion occurs at about 0.06 ps among the ethanol molecules participating in 0 or 1 hydrogen bonds, and at about 0.11 ps among those participating in 2 hydrogen bonds, showing how local hydrogen bonding environment affects not only longer-term reorientation, but also short-term movement.

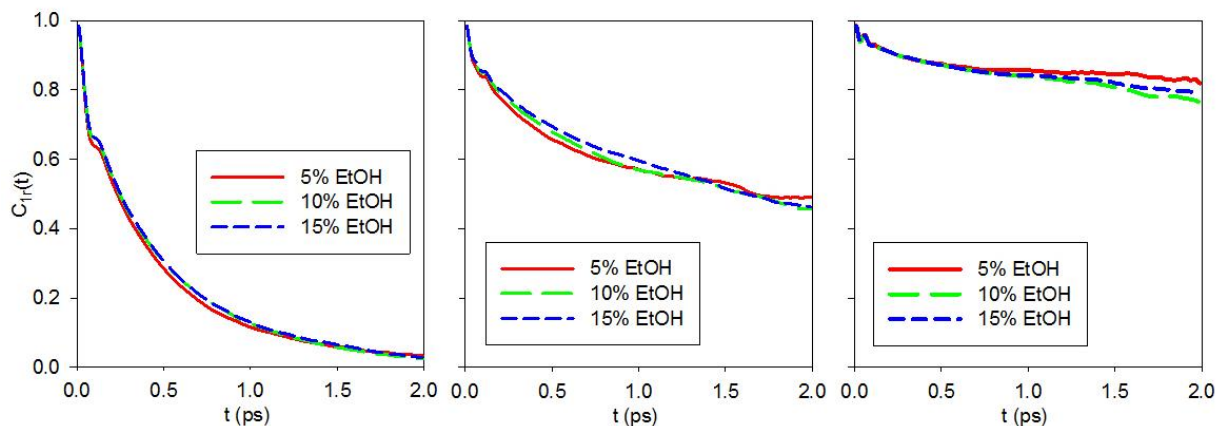


Figure 9.8. Orientational time correlation functions along the O-H vector for populations of ethanol molecules participating in (a) $n=0$, (b) 1, or (c) 2 hydrogen bonds, in the three simulations within the composition series (all at 17-18 MPa and 338 K; 5% ethanol, red; 10% ethanol, green; 15% ethanol, blue).

If the differences in overall orientational dynamics of ethanol molecules between simulations with different ethanol mole fractions arise due to the differing proportions of local hydrogen-bonding environments, we should expect to be able to reproduce the overall dynamics as the weighted average of the orientational TCFs for each individual environment. Figure 9.9 compares this weighted average of the hydrogen bonding subsets (using the data shown Figure 9.8) with the overall orientational TCFs for all ethanol molecules (data shown in Figure 9.6c). Overall agreement is generally good, particularly at short timescales, although divergence begins to occur at longer times as

ethanol molecules change their hydrogen bonding environments and the assumption that overall dynamics reflects the sum of subsets breaks down.

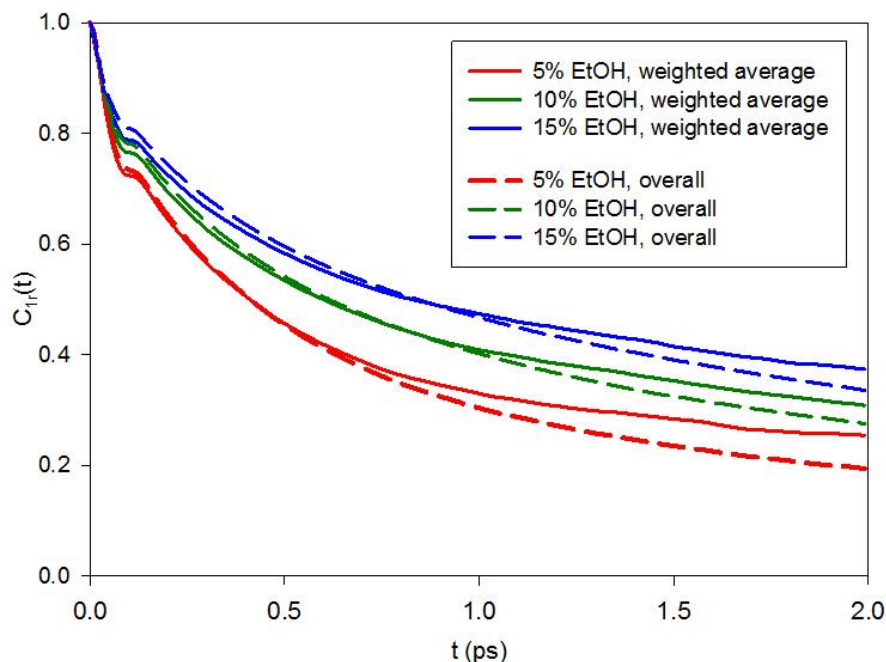


Figure 9.9. Comparison of orientational time correlation functions along the O-H vector for populations of ethanol molecules calculated either as a weighted average of the hydrogen-bonding subsets described in Figure 9.8 or from the overall population of ethanol molecules, as described in Figure 9.6. Data is from composition series of simulations at 338K and 17-18 MPa.

9.3.3 Hydrogen Bond Dynamics

Hydrogen bond time correlation functions describe the timescale on which hydrogen bonds persist; they reflect the fraction of hydrogen bonds still in existence after a given time has elapsed. These correlation functions are described by the function,

$$C_{HB}(t) = \sum_{ij \text{ pairs}} \left\langle \frac{H_{ij}(0)H_{ij}(t)}{H_{ij}(0)} \right\rangle \quad (5)$$

and can describe either *continuous* or *intermittent* hydrogen bonding. For *intermittent* hydrogen bonding time correlation functions, $H_{ij}(t)=1$ if a hydrogen bond exists between

ethanol molecules i and j at time t , and $H_{ij}(t)=0$ otherwise. For *continuous* time correlation functions, $H_{ij}(t)=1$ only if a hydrogen bond existed between molecules i and j at all timesteps sampled between and including 0 and t ; otherwise, $H_{ij}(t)=0$. Continuous hydrogen bond time correlations do not take into account the rapid formation and re-formation of hydrogen bonds that can occur because of vibrational and librational movement of ethanol molecules, and have characteristic lifetimes an order of magnitude or more shorter than those calculated from intermittent hydrogen bond correlations.⁵¹

Hydrogen bond lifetimes – both continuous and intermittent – for each simulation are calculated by integrating the area under the correlation function, using a biexponential fit to account for behavior at time scales beyond those directly calculated from simulation; these lifetimes are reported in Table 9.3. The lifetimes of hydrogen bonds mirror well the trends associated with the numbers and extent of hydrogen bonding – increasing lifetime with increasing ethanol mole fraction and with increasing pressure, decreasing lifetime with increasing temperature.

Furthermore, the lifetime of hydrogen bonds within aggregates of particular sizes can be calculated similarly. Figure 9.10 shows the hydrogen bond TCFs for hydrogen bonds that exist within aggregates of sizes $n=2$ through $n=6$ at $t=0$ for the simulation conducted at 17.2 MPa, 338 K and 10% ethanol mole fraction. (In these calculations, the hydrogen-bonding state or cluster state of ethanol molecules are not held constant, as these would change with the formation or breaking of hydrogen bonds.) The lifespan of hydrogen bonds reaches a maximum within pentamers; aggregates that are both smaller or larger exhibit shorter or longer lifetimes. These trends are consistent across simulation conditions as temperature, pressure, and composition are varied.

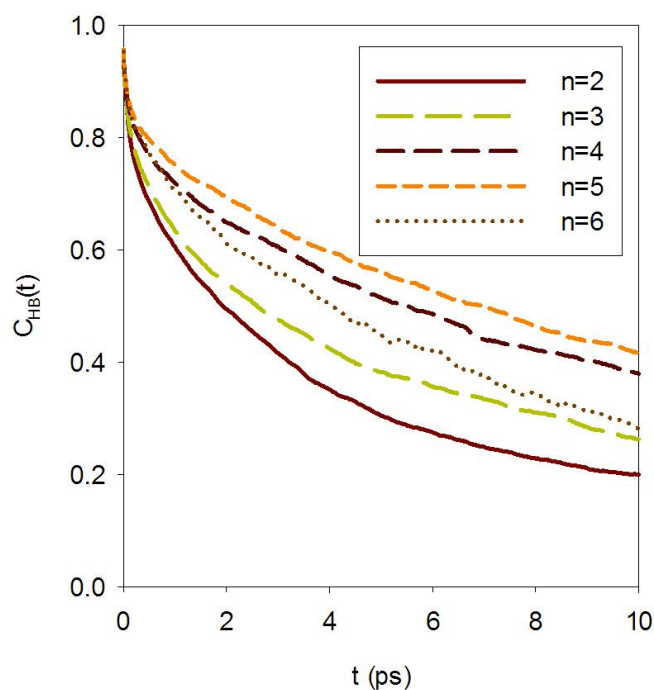


Figure 9.10. Intermittent hydrogen bond time correlations for hydrogen bonds which begin within aggregates of sizes $n=2$ through $n=7$ within the simulation at 17.2 MPa, 338 K and 10% ethanol mole fraction.

The preference for hydrogen-bonded aggregates with four or five ethanol molecules may be attributed in part to particular structures within these aggregates – most likely four or five-membered cyclic structures, as linear hydrogen bonding structures seem unlikely to impart particular stability to clusters of a particular size. This can be validated by examining both the size of the cyclic structures present within larger aggregates, and by examining the lifetime of hydrogen bonds within these cyclic structures. The presence or absence of a cyclic structure within an aggregate can be seen to have an effect on the longevity of hydrogen bonds within its structure. For example, Figure 9.11 shows the non-continuous TCFs for hydrogen bonds within cyclic tetramers and pentamers within the same simulation as Figure 7 (17.2 MPa, 338K, 10% ethanol mole fraction) as compared to hydrogen bonds within non-cyclic tetramers and

pentamers. For both sizes of aggregates, hydrogen bonds within cyclic structures have slightly longer lifetimes (12.4 ps for tetramers, 12.1 ps for pentamers) than those within non-cyclic aggregates of the same size (10.3 ps for tetramers, 11.0 ps for pentamers). This adds further weight to the hypothesis that cyclic structures impart greater stability to ethanol aggregates, modestly lengthening hydrogen bond lifetimes and constraining the translational and rotational dynamics of ethanol within the mixture.

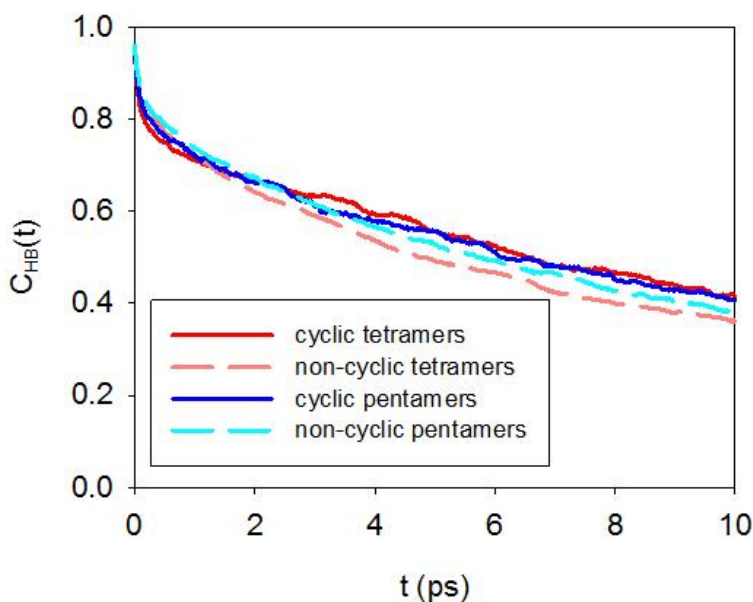


Figure 9.11. Intermittent hydrogen bond time correlation functions for hydrogen bonds within tetramers and pentamers with and without cyclic structures in the system at 10% ethanol mole fraction, 338K, 17.2 MPa.

Hydrogen bond TCFs can also be used to probe the dynamics of ethanol aggregates. At each point in time, an ethanol molecule belongs to an aggregate of size n , and the breaking or formation of a hydrogen bond – unless it simply disrupts or creates a cyclic structure within an aggregate – will result in a change in aggregate size. One can therefore examine the dynamics of these clusters by means of a cluster size time correlation function, C_{CL} ⁴⁹ depicting the life span of a cluster a given size – the fraction

of clusters of size n that remain that size after a given time t has elapsed. Figure 9.12 reports cluster size time correlation functions for the simulation at 338K and 17.2 MPa with 10% ethanol mole fraction. These cluster size TCFs show faster decay than the hydrogen bond TCFs associated with each cluster size n , shown in Figure 9.10; this is expected, as cluster size can be modified not only by the breaking of multiple different hydrogen bonds within a cluster, but also by the addition of a new hydrogen-bonded ethanol molecule to that cluster. For large clusters, the disruption of any one out of many hydrogen bonds would alter cluster size. It is therefore notable that aggregates of size $n=2$, and 4, 5 clearly exhibit similarly long lifetimes, consistent with their prevalence in the distribution of aggregate sizes; the lifespans of both trimers and larger aggregates are consistently shorter. This result precisely parallels that observed by examining hydrogen bond TCFs directly.

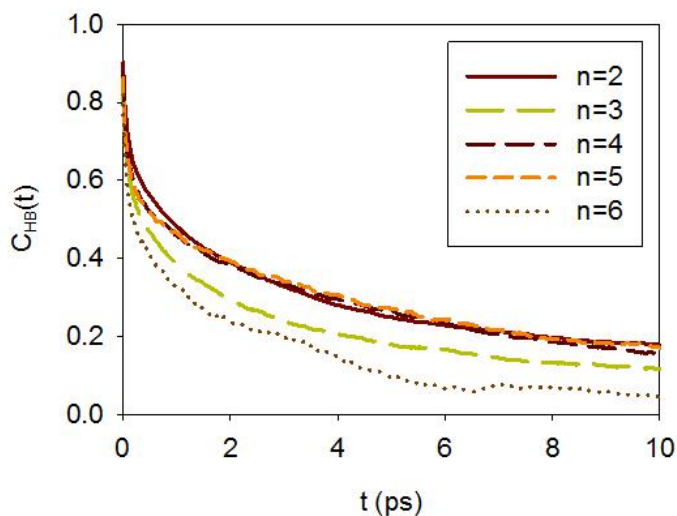


Figure 9.12. Cluster size time correlation functions for aggregates of size n , among the mixtures at 17.2 MPa, 338K and containing 10% ethanol mole fraction.

9.4 Discussion

We have examined both the structure and dynamics of sc-CO₂/ethanol mixtures under a range of thermodynamic conditions, with particular attention to hydrogen-bonded ethanol aggregates. Only slight changes in structure were seen with increasing pressure, while with decreasing temperature or increasing ethanol mole fraction, a greater degree of ethanol-ethanol association and greater aggregate size was noted. Similarly, hydrogen bonding lifetimes and rotational and translational dynamics paralleled these trends.

Molecular dynamics simulations offered the opportunity to explore the relationship between structure and dynamics by examining the behaviors of particular subsets of ethanol molecules that exist in certain environments – particularly to help understand differences in dynamics with differing ethanol compositions. Reorientation times for ethanol molecules in similar hydrogen-bonding environments (at comparable temperatures and pressures) are nearly identical, demonstrating that increasing composition, and therefore increasing the number of ethanol molecules in more constrained environments, is the determining factor in the overall dynamical trends. Similarly, differences in average diffusion coefficients for ethanol molecules in different simulated can be explained by the shifts in ethanol environments – in particular, the greater number of neighboring molecules with which hydrogen bonding occurs.

Among hydrogen-bonded aggregates, tetramers and pentamers were also noted to be unusually prevalent, presumably stabilized in part by the presence of cyclic structures within clusters of this size. Examination of hydrogen-bond dynamics within clusters showed that hydrogen bond lifetimes were longest within these clusters, and, in fact, were longer amongst those tetramers and pentamers containing cyclic structures.

Chapter 10: Intermolecular Interactions Around Solutes in Supercritical Carbon Dioxide/Ethanol Mixtures

As described in the previous chapter, although supercritical CO₂ is an attractive solvent for analytical and industrial processes, polar organic molecules tend to have poor solubility in this non-polar substance. The addition of polar cosolvents, such as methanol, ethanol, or acetone, has been seen substantially enhance the solubility of organic solutes.⁵² These cosolvents are often called “entrainers,” reflecting that they are believed to preferentially surround the solute molecules and “carry them along” as they move through the solution. This description suggests relationships between the behavior of entrained polar solutes and the behavior of solvent molecules, including local composition enhancements of one component of the cosolvent mixture around the solute – and such composition enhancements have been detected spectroscopically.^{6b, 9b,}

35, 53

As has been discussed in the previous chapter, just as measurable properties such as pressure and viscosity change markedly in supercritical systems with relatively small changes in temperature and pressure, similarly, in cosolvent systems, properties such as solubility and solvent mobility change substantially with modest changes in solvent mixture composition.^{20, 22} Recent work by Suarez-Iglesias has shown pronounced differences in solute mobility in CO₂/ethanol mixtures as thermodynamic conditions are varied.³⁴

Although not all of the molecular-level behaviors that can be calculated from the coordinates generated in molecular dynamics simulations are easily accessible from

experiment, the diffusion coefficients of solutes are readily accessible and have been reported for several solutes in a range of thermodynamic conditions. These results therefore serve as a target system against which simulations can be compared, and motivated the studies described in this chapter, in which we have simulated these three-component systems (one solute and two cosolvents) in order to probe the molecular-level origins of these mixtures' behavior. In this chapter, we attempt to use these simulations of these three-component systems to make contact with and help explain the molecular-level origin of the behavior that manifests in experimental measurements of these and similar systems.

Building upon our understanding of the structure and dynamics of CO₂-ethanol systems described in the previous chapter, we are particularly interested in the local environment around solutes. What does this local environment look like – is it similar or different in composition from the bulk solute mixture? Does this differ for solutes that are different in size or polarity or for solute mixtures of different compositions? For larger solutes, how does the solvent structure vary around different sites in the molecule? Do we observe evidence of the “entraining” effect described above? How does local structure change with time, and what effect does this local structure have on the dynamics of the solute molecules?

10.1 Simulation Parameters and Conditions

In this chapter, we describe the analysis of four molecular dynamics simulations, each containing one solute molecule and 500 molecules (a mix of carbon dioxide and ethanol) that comprise the solvent mixture.

To make contact with work by Suarez-Iglesias et al, the solute molecules chosen for simulation were benzyl acetate and dibenzyl ether. These two solutes, both studied experimentally at a range of temperatures, pressures and compositions, differ significantly in their molecular weight (dibenzyl ether being substantially more massive) and, potentially, in their hydrogen-bonding capacity because of their differing functionalities.

Carbon dioxide and ethanol were simulated using the same potential functions as described in the previous chapter – ethanol with OPLS-UA⁴¹ and carbon dioxide with the EPM2 models⁴⁰ – as described in the previous chapter. Solutes were simulated by TRAPPE-UA potentials.⁵⁴ These models themselves in fact borrow significantly from the OPLS-UA potential functions for treatment of ether and ketone groups, suggesting a reasonable compatibility with the ethanol sites that they are expected to interact most strongly with.

In these united-atom potential functions, hydrogens are not represented explicitly, but are considered part of the sites centered on their location of the adjacent carbon atom. Each of the carbon and oxygen sites within solutes are each modeled as a Lennard-Jones site combined (in some cases) with point charges. Lennard-Jones parameters and charges for each solute are given in Table 10.1. Unlike sites were treated using standard Lorentz-Bertholet combining rules.

All bond lengths were fixed at the values given in Table 10.2, as was the hexagonal geometry of the phenyl ring. However, the remaining flexible bond angles are modeled using harmonic potentials of the form:

$$U(\theta) = \frac{k}{2}(\theta - \theta_0)^2 \quad (1)$$

where θ_0 is the equilibrium bond angle, and k is the force constant, specified for each flexible bond angle in Table 10.3. The table also specifies θ_0 values for fixed bond angles.

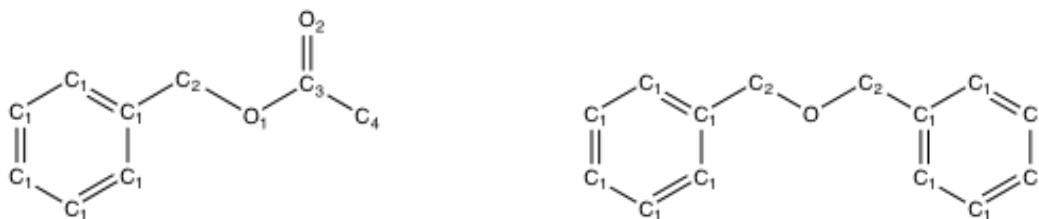


Figure 10.1. Labeled types of sites, as described within Tables 10.1-10.4, within solute molecules – benzyl acetate (left) and dibenzyl ether (right).

Table 10.1. Lennard-Jones parameters and charges for sites within solute molecules. Site indices are shown in Figure 10.1.

Site	σ (Å)	ϵ (kcal/mol)	q (e)
Benzyl acetate: C ₁	3.695	0.1103	0
Benzyl acetate: C ₂	3.950	0.0914	+0.250
Benzyl acetate: O ₁	2.800	0.1093	-0.400
Benzyl acetate: C ₃	3.900	0.0815	+0.550
Benzyl acetate: O ₂	3.050	0.1570	-0.450
Benzyl acetate: C ₄	3.750	0.1947	+0.050
Dibenzyl ether: C ₁	3.695	0.1103	0.000
Dibenzyl ether: C ₂	3.750	0.1947	+0.250
Dibenzyl ether: O	2.800	0.1093	-0.500

Table 10.2. Fixed bond lengths within solute molecules.

Bond pair	Bond length (Å)
Benzyl acetate: C ₁ -C ₁	1.40
Benzyl acetate: C ₁ -C ₂	1.56
Benzyl acetate: C ₂ -O ₁	1.344
Benzyl acetate: C ₃ -O ₁	1.344
Benzyl acetate: C ₃ -O ₂	1.20
Benzyl acetate: C ₃ -C ₄	1.52
Dibenzyl ether: C ₁ -C ₁	1.40
Dibenzyl ether: C ₁ -C ₂	1.56
Dibenzyl ether: C ₂ -O	1.41

Table 10.3. Bond angles and, where applicable, force constants for harmonic bond angle potentials within for the two solute molecules.

Bond angle sites	Bond Angle (θ)	k (kcal/mol)
Benzyl acetate: C ₁ -C ₁ -C ₁	120	fixed
Benzyl acetate: C ₁ -C ₁ -C ₂	120	fixed
Benzyl acetate: C ₁ -C ₂ -O	110	140.29
Benzyl acetate: C ₂ -O ₁ -C ₃	115	124.20
Benzyl acetate: O ₁ -C ₃ -O ₂	110	140.29
Benzyl acetate: O ₁ -C ₃ -C ₄	125	124.20
Dibenzyl ether: C ₁ -C ₁ -C ₁	120	fixed
Dibenzyl ether: C ₁ -C ₁ -C ₂	120	fixed
Dibenzyl ether: C ₁ -C ₂ -O	112	99.956
Dibenzyl ether: C ₂ -O-C ₂	112	120.027

Table 10.4. Coefficients for the dihedral angle potentials (in equation (2) for the non-fixed torsional potentials in solute molecules.

Dihedral angle sites	U_0 (kcal/mol)	U_1 (kcal/mol)	U_2 (kcal/mol)	U_3 (kcal/mol)
Benzyl acetate: C ₁ -C ₁ -C ₂ -O ₁	0	0.3510	-0.1060	1.5300
Benzyl acetate: C ₂ -O ₁ -C ₃ -O ₂	0	1.4414	0.3254	1.1093
Benzyl acetate: C ₂ -O ₁ -C ₃ -O ₄	0	4.2884	4.1691	0.3921
Dibenzyl ether: C ₁ -C ₁ -C ₂ -O	0	0.3510	-0.1060	1.5300
Dibenzyl ether: C ₁ -C ₂ -O-C ₂	0	1.4414	-0.3254	1.1093

Torsional potentials outside the fixed benzene ring are all modeled using the triple cosine function:

$$U(\varphi) = U_0 + \frac{1}{2}U_1(1 + \cos \varphi) + \frac{1}{2}U_2(1 + \cos 2\varphi) + \frac{1}{2}U_3(1 + \cos 3\varphi) \quad (2)$$

where the coefficients U_n for flexible dihedral angles are specified in Table 10.4.

Simulations were equilibrated first for 500 ps in a NPT ensemble at a temperature of 333 K and 25 MPa in order to establish an appropriate simulation volume. A representative configuration from this equilibration run at the average box size was taken as the starting point for a NVT production run, which was equilibrated for 100 ps and then continued for a production run of 2.4 ns from which simulation data is drawn. As can be seen in Table 10.5, although the production runs for the dilute ethanol systems resulted in calculated pressures comparable to the intended pressure of 25 MPa, however, the “concentrated” ethanol systems yielded substantially lower calculated pressures. Caution should therefore be taken in interpreting differences

between simulations as due solely to differences in composition – however, based on the results reported in Chapter 2, differences in pressure in this range are seen to have relatively small differences in solution structure or dynamics, so it is likely that contrasts between the “dilute” and “concentrated” ethanol systems are, in fact, primarily due to composition differences.

Table 10.5. System parameters for the four simulations discussed in this chapter.

	Number of CO₂ molecules	Number of EtOH molecules	Simulation density (g/cm³)	Calculated pressure (MPa)
Benzyl acetate in “dilute” EtOH	476	24	0.828	22.2
Benzyl acetate in “concentrated” EtOH	382	118	0.817	14.0
Dibenzyl ether in “dilute” EtOH	476	24	0.846	25.7
Dibenzyl ether in “concentrated” EtOH	382	118	0.814	13.4

10.2 Pair Correlation Functions

Radial distribution functions (RDFs) were calculated for each type of site in solute molecules, and are shown in Figures 10.1-10.3, as graphs of the pair correlation functions for the CO₂ sites and ethanol sites separately. These figures clearly show different organizational structures around different solute sites.

Around the carbons in each solute’s phenyl ring, there is relatively little evidence of short-range structure. Figures 10.1a and 10.3a illustrate the effects of CO₂’s linearity on radial distribution functions, with one oxygen peak (the nearest “end” of a CO₂ molecule) around 3.5 Å, a carbon peak around 4.5 Å, and another oxygen peak around

5.5 Å. Similarly, only a small degree of ordering around the benzyl carbons can be seen in the ethanol RDFs in Figures 10.1b and 10.3b – although the carbon sites generally lie closer to the phenyl ring than the oxygen and hydrogen sites, there is no strong ordering indicating a common orientation with respect to these carbons. Furthermore, none of these peaks rise above 1.3, indicating only a mildly favorable interaction with either solvent molecule.

This contrasts markedly with the most negatively charged sites in each molecule – the ether oxygen in dibenzyl ether (Figures 10.3e and 10.3f), and the carbonyl oxygen in benzyl acetate (Figures 10.2c, 10.2d). These RDFs exhibit no clear structuring of the carbon dioxide molecules besides their presence within the first solvation shell, but pronounced ordering of the sites within the ethanol molecule, with the nearest hydrogen peak sitting less than 2 Å from the oxygen (rising to a maximum above 6), and the alcohol's oxygen peak at less than 3 Å. Note the similarity between the pair correlation functions for ethanol sites for the benzyl acetate's carbonyl oxygen and the dibenzyl ether's ether oxygen. Each shows clear and pronounced peaks for each of the four sites within the ethanol molecule in sequence, with the sharpest and highest peak for hydrogen at about 2 Å. This is a strong indication of hydrogen bonding around these sites. The ether oxygen in benzyl acetate, however, does not show structure that is nearly as well-defined, although the hydrogen peak at about 2 Å indicates that a smaller degree of hydrogen bonding may be occurring around this negatively charged site. Sites adjacent to hydrogen-bonding sites – such as the benzyl carbons in dibenzyl ether (Figures 10.3c and 10.3d), or the carbonyl carbon in benzyl acetate (Figures 10.2a and 10.2b), also show clear geometrical ordering with respect to those sites, although not to

the same extent; this can be attributed to a “spillover” effect of the hydrogen bonding activity close to, but not involving, these particular sites.

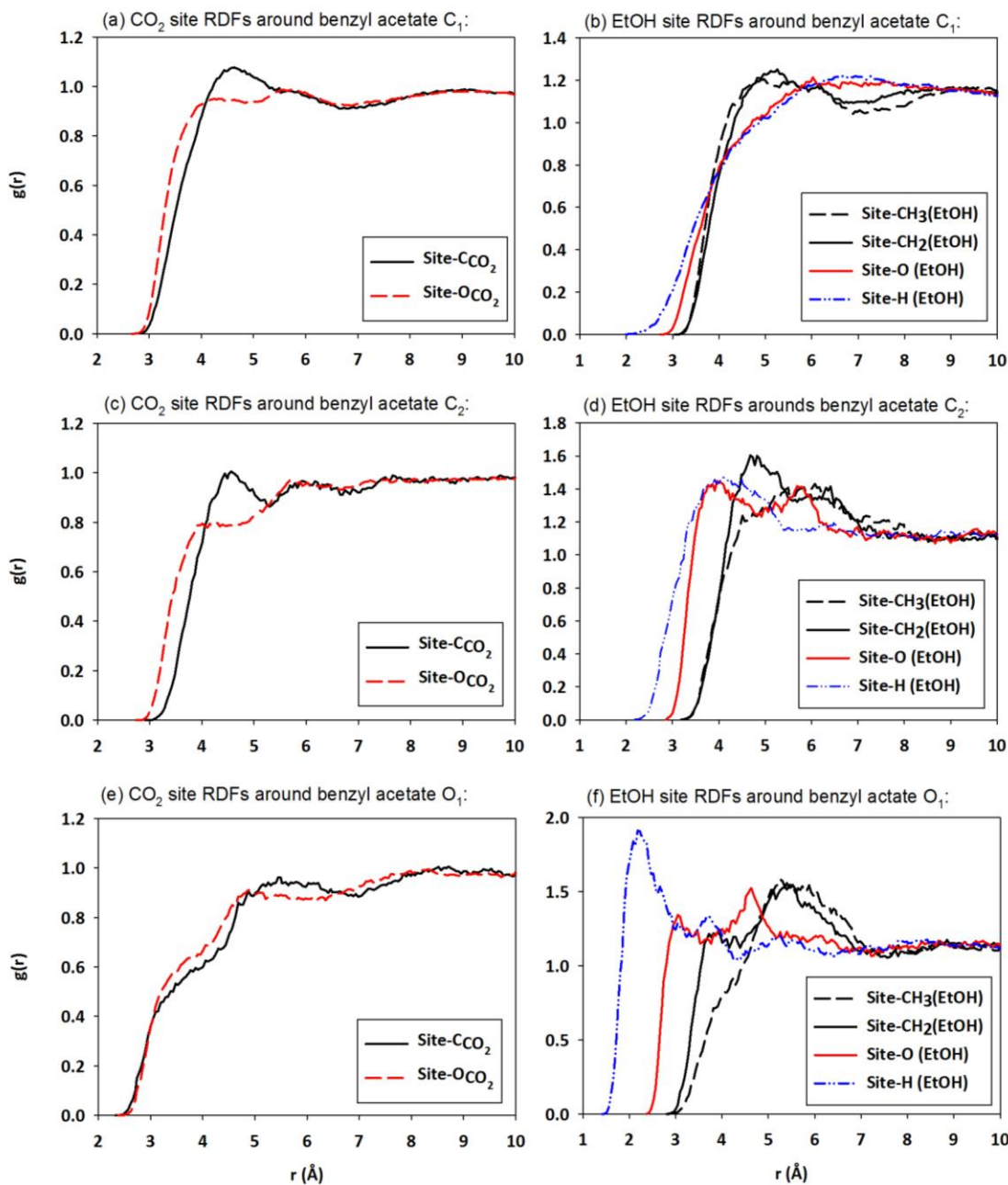


Figure 10.2. Radial distribution functions for carbon dioxide sites (left-hand column) and ethanol sites (right-hand column) around sites on simulated benzyl acetate molecules – around phenyl ring carbons, C_1 (first row); the benzyl carbon, C_2 (second row), and the ether oxygen O_1 (third row).

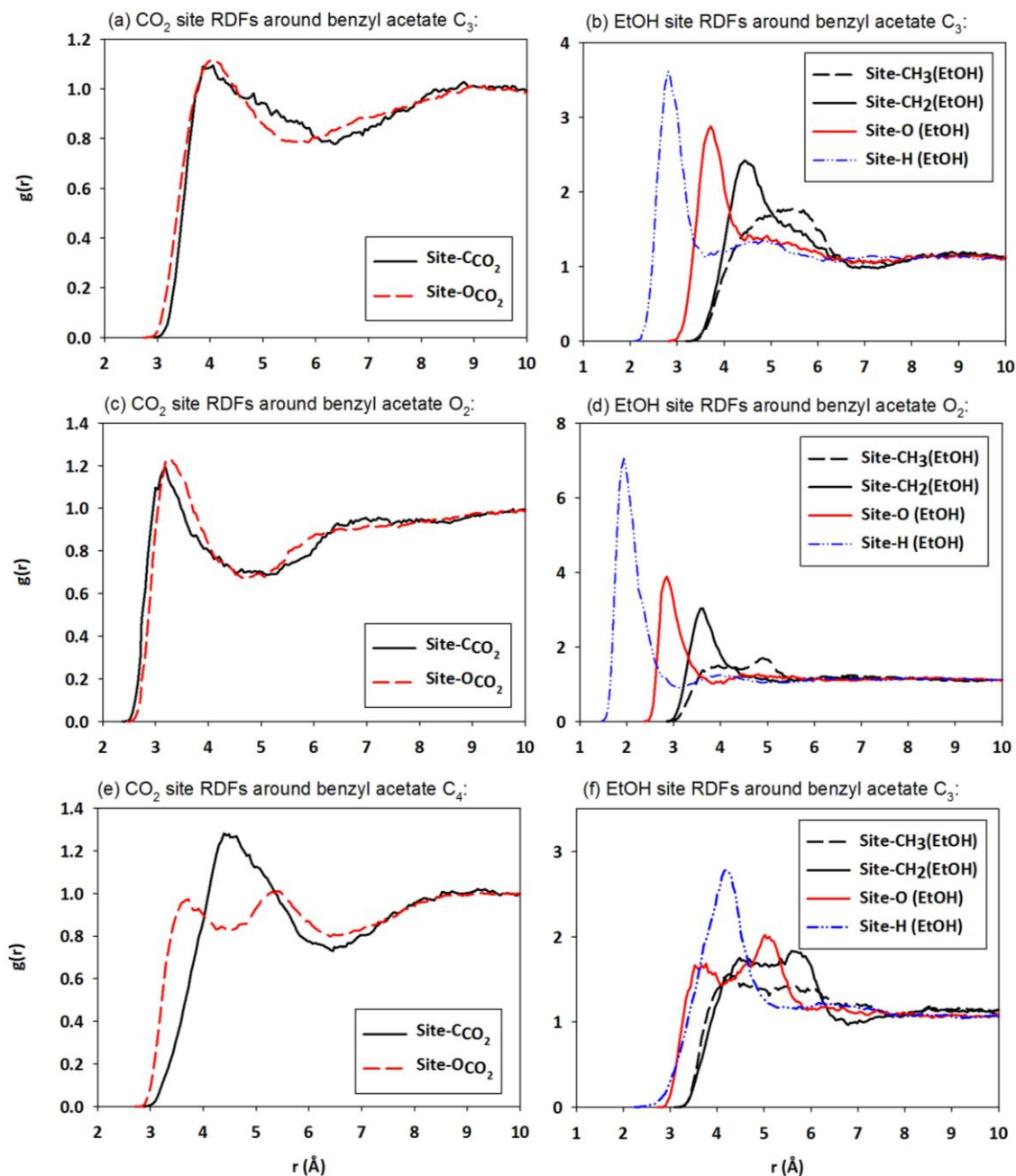


Figure 10.3. Radial distribution functions for carbon dioxide sites (left-hand column) and ethanol sites (right-hand column) around sites on simulated benzyl acetate molecules – around the carbonyl carbon, C₃ (first row); the carbonyl oxygen, O₂ (second row), and the terminal carbon, C₄ (third row).

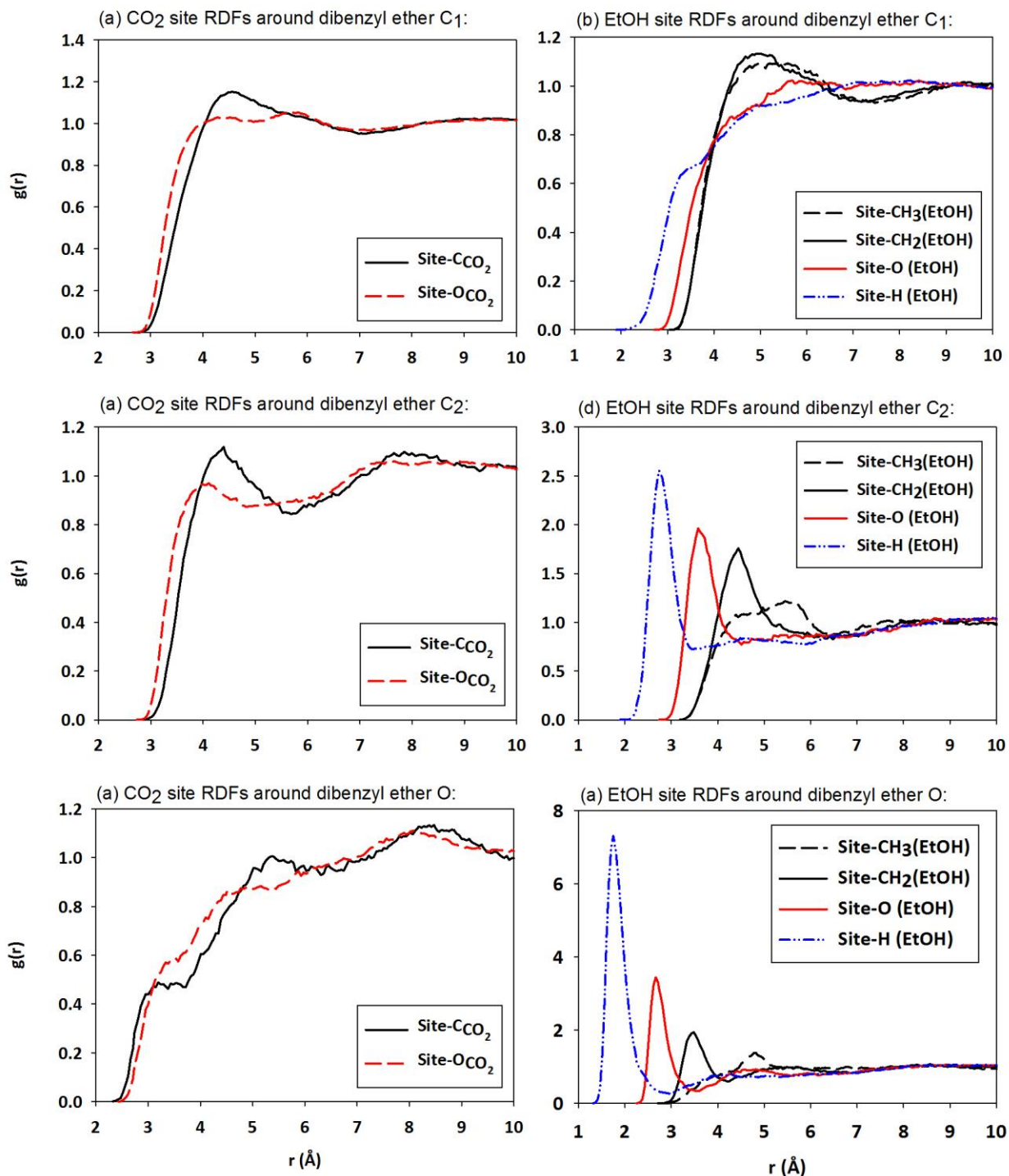


Figure 10.4. Radial distribution functions for carbon dioxide sites (left-hand column) and ethanol sites (right-hand column) around sites on simulated dibenzyl ether molecules – around the benzyl carbon, C₁ (first row); the benzyl carbon, C₂ (second row), and the ether oxygen (third row).

As in Chapter 2, local mole fractions around each solute site were calculated. These plots illustrate the proportion of solvent molecules of each type at a given distance from the solute site, allowing us to compare the mole fraction with the bulk mole fraction (which the local mole fraction should be equal to at long distances). Local mole fraction plots shows marked local composition enhancement around the carbonyl oxygen in benzyl acetate (Figure 10.5e) – doubling the concentration of ethanol in the first solvation shell in the more concentrated solution to 51%, and increasing the local mole fraction in the dilute solution to 22% -- more than a five-fold increase over the bulk concentration. Interestingly, this composition enhancement drops off rather markedly after about 4 Å, where the composition returns to close to the bulk density. Note that the local mole fraction plots do not indicate a composition enhancement in the neighborhood of the phenyl ring carbons; around this portion of each solute, composition essentially mirrors the bulk composition.

10.3 First Solvation Shell

Based on pair correlation functions, we have defined a first solvation shell as corresponding to the presence of a solvent molecule site within 5.5 Å of a solute molecule site. Within the preponderance of the local mole fraction plots, this cutoff includes the first peaks corresponding to each site for all of the sites within the nearest molecule. Furthermore, based on the local mole fraction plots, this cutoff fully encompasses the first peak associated with the nearest neighboring solute molecules and the majority of the area that deviates from bulk solute composition.

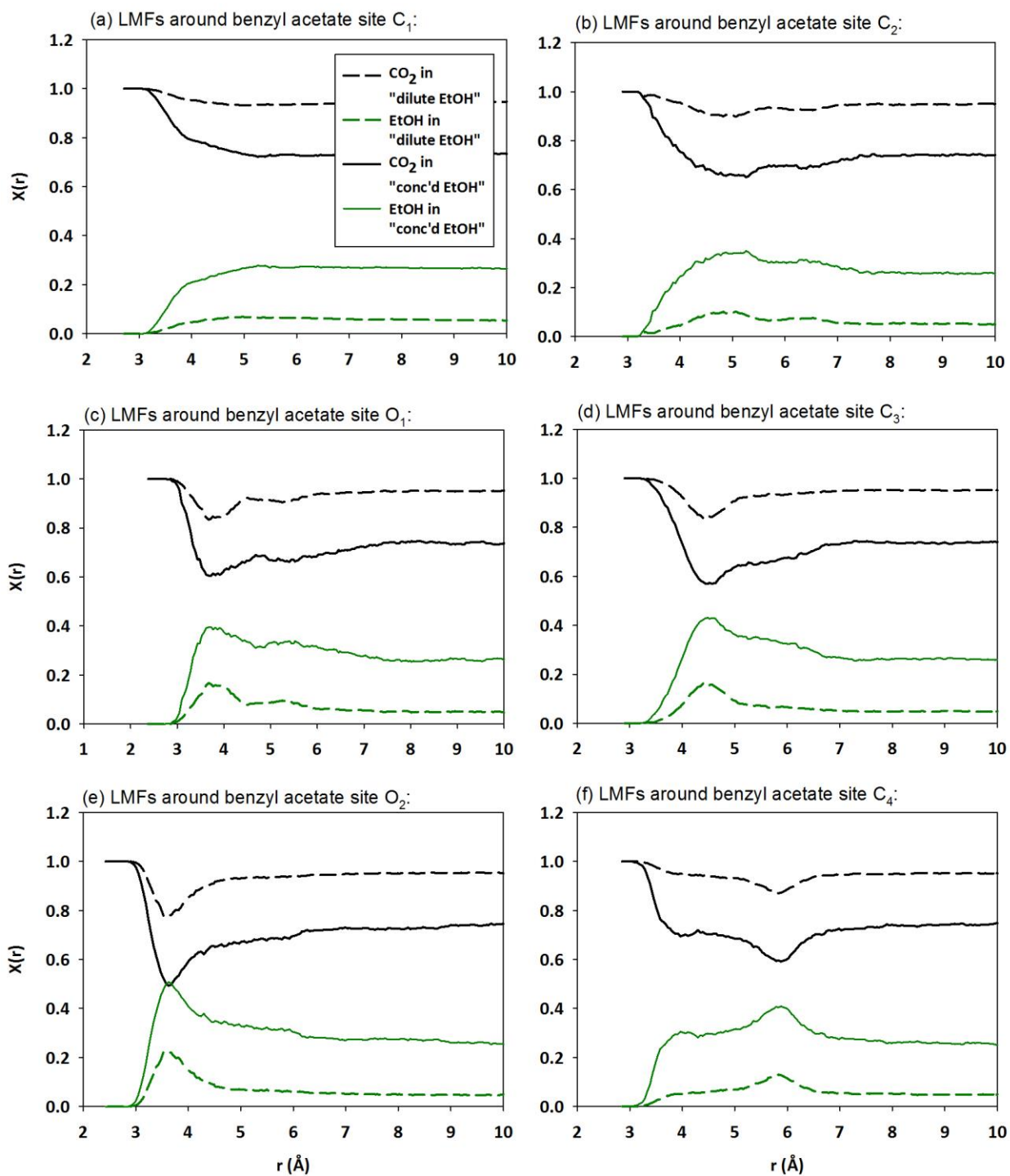


Figure 10.5. Local mole fractions (LMFs) based on solute center of mass around sites on simulated benzyl acetate molecules for both “dilute” ethanol (4.8% mole fraction) and “concentrated” ethanol (23.6%) solvent mixtures.

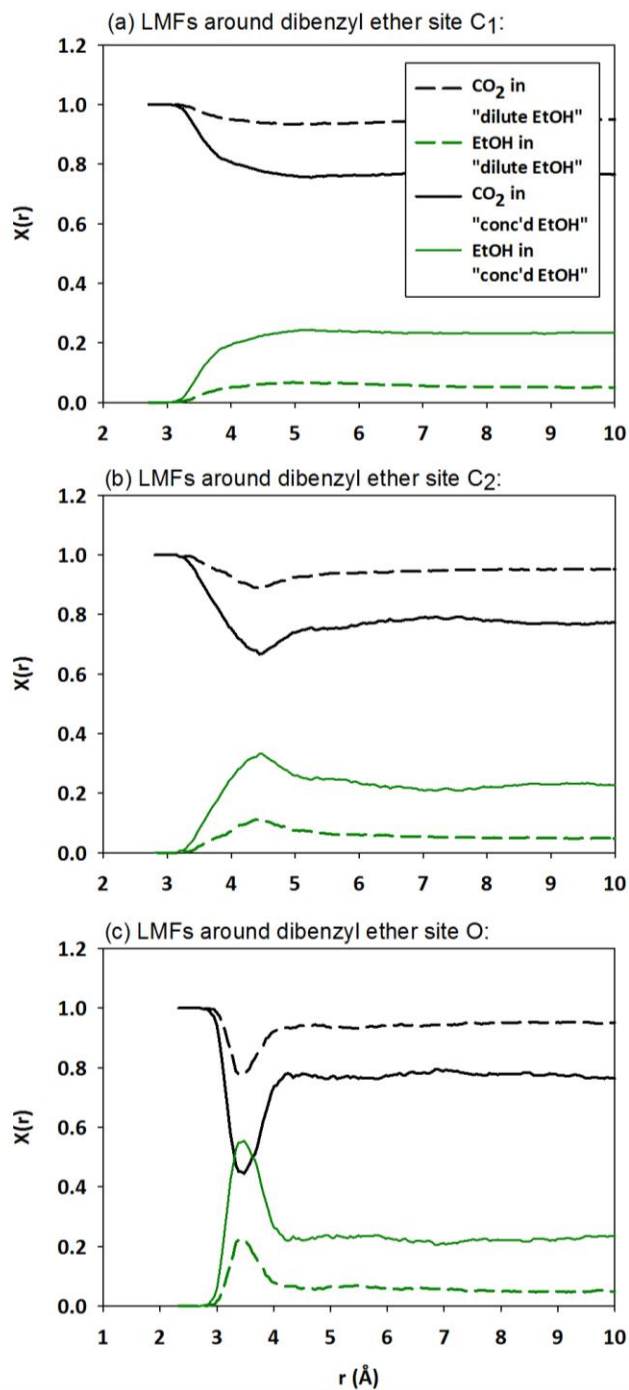


Figure 10.6. Local mole fractions (LMFs) based on solute center of mass around sites on simulated dibenzyl ether molecules for both “dilute” ethanol (4.8% mole fraction) and “concentrated” ethanol (23.6%) solvent mixtures.

A depiction of a dibenzyl ether molecule and its first solvation shell is shown in Figure 10.7. Interestingly, this illustration shows two ethanol molecules engaged in hydrogen bonding with the negatively charged ether oxygen within the solute molecules. At the bottom-right, three ethanol molecules can be seen forming a hydrogen-bonded cluster.

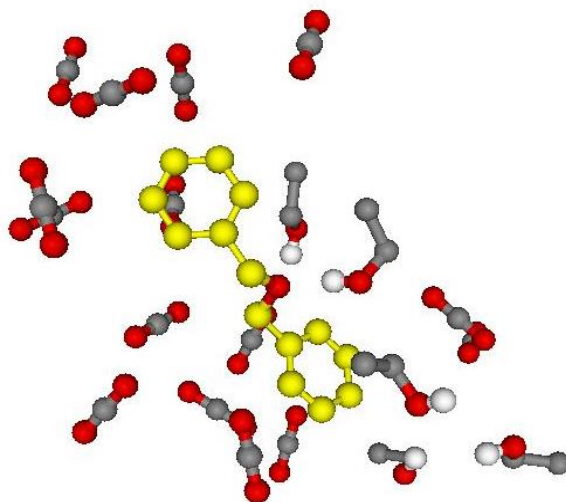


Figure 10.7. Snapshot of molecular configuration for a dibenzyl ether molecule and solvent molecules (CO_2 and ethanol) in the first solvation shell.

Having defined the first solvation shell, the average number of each type of solvent molecule that extends within the first solvation shell – are calculated (below). For benzyl ether, the total number of molecules in the first solvation shell is about 17, whereas the larger dibenzyl ether molecules are surrounded by about 21 molecules. Overall, the composition of the first solvation shell does not differ substantially from that of the bulk mixture – ethanol mole fraction is only enhanced by 2-4% over bulk values in each case. However, as seen in the radial distribution functions above,

preferential interactions between ethanol and the solute are limited to one or two sites on the solute molecule. These locations where ethanol composition enhancements are seen encompass only a small portion of the total volume of the first solvation shell.

Table 10.6. First solvation shell coordination numbers and residence lifetimes around solutes in each simulation.

Solute	Bulk EtOH mole fraction	Coordination number			Residence lifetimes	
		CO ₂	EtOH (mole fraction w/in 1st shell)	Total	CO ₂	EtOH
Benzyl acetate	4.8%	16.4	1.1 (6%)	17.5	6.1 ps	10.5 ps
	23.6%	12.5	4.6 (27%)	17.1	6.4 ps	11.7 ps
Dibenzyl ether	4.8%	19.6	1.7 (8%)	21.3	6.5 ps	9.2 ps
	23.6%	15.0	5.7 (28%)	20.7	7.1 ps	12.6 ps

10.4 Solvation Shell Residence Times

The “entraining” description of the role of cosolvents suggests that a solute and a polar cosolvent travel along together to some extent as they diffuse through the solution. If that is the case, one would expect to see different residence times within the first solvation shell for solutes with stronger interactions with ethanol.

An associated residence time autocorrelation function can be defined under equation 3:

$$C_{ssh}(t) = \frac{\langle s(0) \cdot s(t) \rangle}{\langle s(0) \rangle} \quad (3)$$

where $s(t) = 1$ if a solvent molecule lies in the first solvation shell at time t , and $s(t)=0$ otherwise. Residence time autocorrelation functions for each molecule in each simulation are shown in Figure 10.8, and associated lifetimes, calculated by fitting the curve to a dual exponential fit, are reported in Table 10.6.

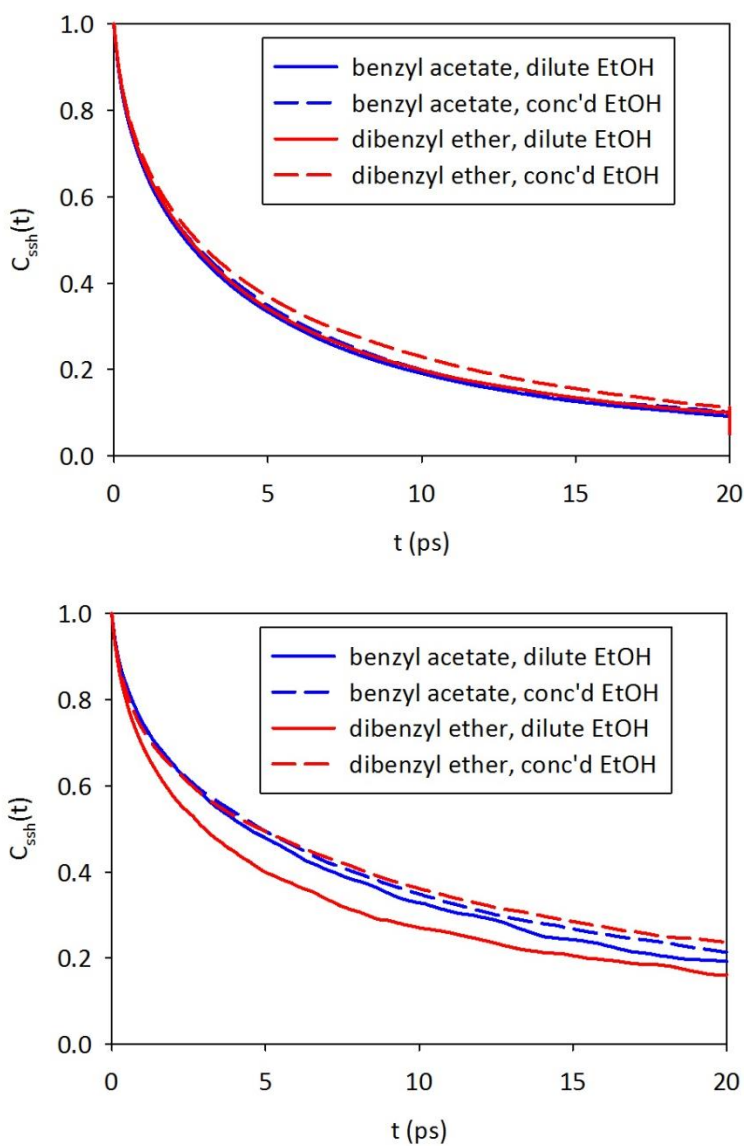


Figure 10.8. First solvation shell residence autocorrelations for CO_2 (top) and EtOH (bottom) in all 4 systems.

In every simulation, CO₂ has a much shorter lifetime in the first solvation shell than ethanol, just as CO₂ molecules were seen in Chapter 2 to have more rapid rates of diffusion. For CO₂, longer time-scale movement of solvent molecules into or out of the first solvation shell appears to be governed as much by bulk solvent composition – residence lifetimes become 5-10% longer as solvent composition is increased, as by the identity of the solute (which affects solvation shell lifetime by 7-11% within each composition condition). For ethanol, the effects of both solvent mixture composition and solute identity are more pronounced; in both cases, movement of ethanol into and out of the first solvation shell is slowed in the simulations with greater ethanol mole fraction; however, it is around dibenzyl ether that this effect is most pronounced. This may be due to the differing nature of hydrogen bonding interactions between ethanol and each of the two different solutes studied.

10.5 Ethanol/Solute Hydrogen Bonding

As it appears that it is hydrogen-bonding interactions which lead to local enhancements in ethanol mole fraction in the first solvation shell and therefore to longer residence times in that region, we then chose to examine hydrogen-bonding interactions specifically. As in Chapter 2, the coordinates generated from simulation trajectories can be analyzed for arrangements consisting with hydrogen bonding. The same standard geometric criterion were applied: an intermolecular O-O distance less than 3.5 Å, an intermolecular O-H distance less than 2.6 Å, and an O---O-H angle deviating from linearity by less than 30 degrees.¹⁸⁻¹⁹

Based on this criterion, the fraction of simulation configurations in which particular solute sites participate in hydrogen bonds with ethanol can be calculated; these are reported in Table 10.7. As expected, it is the carbonyl oxygen in benzyl acetate which is the most likely to be a part of a hydrogen bond at any given time. This probability increases with increased ethanol mole fraction in the solvent mixture. Similarly, hydrogen bond lifetimes, calculated based on continuous hydrogen bond autocorrelation functions (see Chapter 2) are longer for the carbonyl oxygen than for the ether oxygen in benzyl acetate. These lifetimes also remain similar with changes in solvent mixture composition. However, although ethanol molecules are less likely to engage in hydrogen bonding with the ether oxygen in dibenzyl ether than with the ether oxygen in benzyl acetate, the hydrogen bonds that do form are longer-lived. There are therefore somewhat conflicting trends in hydrogen-bonding behavior – while benzyl ether possess more hydrogen-bonding sites and a greater likelihood of participating in hydrogen bonding, hydrogen bonds with dibenzyl ether are actually longer-lived.

Table 10.7. Hydrogen bond lifetimes and fractions of time in which each solute oxygen site in each simulation participate in hydrogen bonding, as defined by a geometric criterion.

Solute – site		EtOH mole fraction	H-bond lifetime	Percent of time H-bonded
Benzyl acetate	Ether oxygen	4.8%	0.18 ps	7.1%
	Carbonyl oxygen	4.8%	0.38 ps	21.4%
	Ether oxygen	23.6%	0.19 ps	14.2%
	Carbonyl oxygen	23.6%	0.41 ps	45.5%
Dibenzyl ether	Ether oxygen	4.8%	0.56 ps	8.9%
	Ether oxygen	23.6%	1.23 ps	41.3%

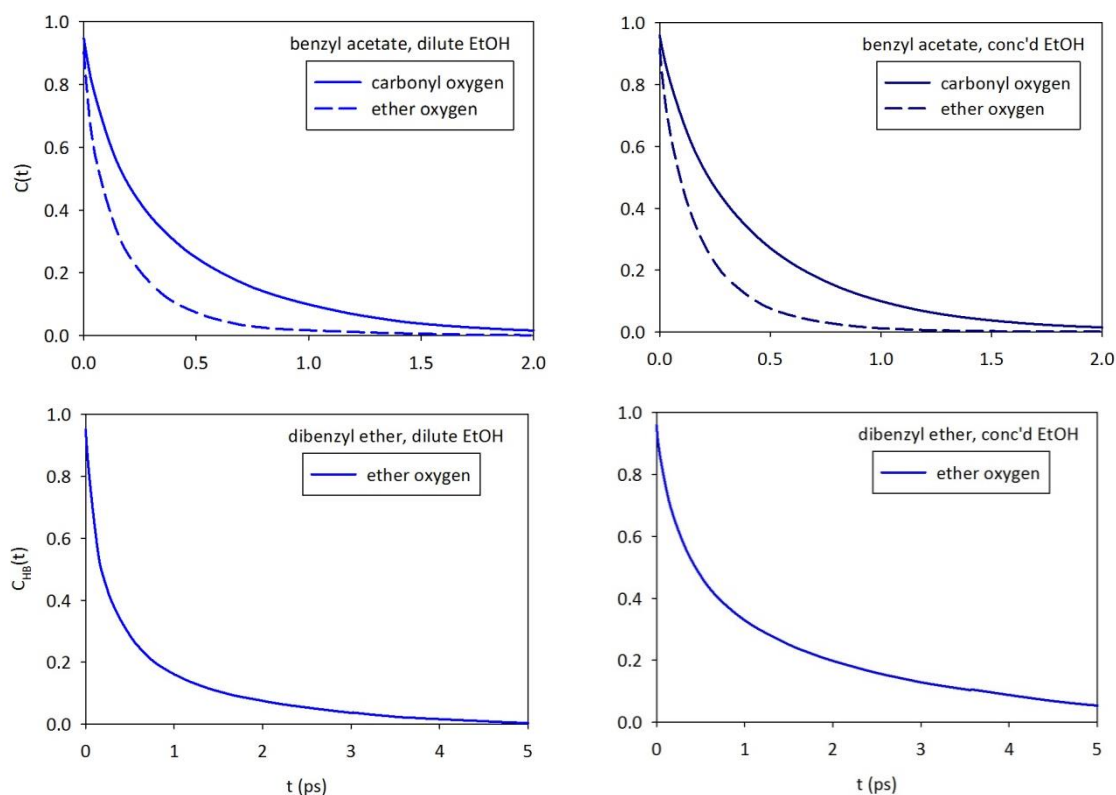


Figure 10.9. Continuous hydrogen bond time autocorrelation functions for each oxygen site in solute molecules, for each simulation.

10.6 Diffusion

Mean squared displacements calculated from simulation trajectories based on the center of mass of each component are shown in Figure 10.10. In every simulation, CO_2 shows the fastest diffusion, while more strongly self-associating ethanol molecules diffuse more slowly. As the ethanol mole fraction is increased in each system, the diffusion of all components are slowed – ethanol substantially, by about 30%, while the diffusion of carbon dioxide is only slowed by less than 20%. These variations in the diffusion of cosolvent components are consistent with the data from systems without solutes shown in the previous chapter.

Ethanol mole fraction has a substantial effect on solute mobility as well. The larger solute, with a larger hydrodynamic radius, diffuses more slowly than either solvent species, but diffusion for both solutes slows substantially with increasing ethanol mole fraction, bringing the solute's diffusion rates close to those of ethanol molecules. This is consistent with the markedly greater ethanol mole fraction within the first solvation shell, as ethanol molecules form hydrogen-bonded complexes with solute molecules.

Diffusion coefficients can be calculated for each component in each mixture, as the limiting slope of mean squared displacement with respect to time can be related to the diffusion coefficient via the Einstein relation, as described in Section 9.3.1, and diffusion coefficients for solutes can be compared to the values determined experimentally by Suarez-Iglesias et al.³⁴ Simulated diffusion coefficients for the solutes in the dilute ethanol mixtures were within 15% of those determined experimentally; in the concentrated ethanol mixtures, solute diffusion coefficients were somewhat (12-33%) greater than experimental values. These differences – particularly those in the concentrated ethanol conditions – suggest that these simulations may not fully account for the influence of ethanol or ethanol aggregates on the motion of solute molecules.

In contrast with experimental measurements, diffusion of benzyl acetate is calculated to be slower than for dibenzyl ether under both composition conditions. This is unexpected, as dibenzyl ether is the larger of the two solute species and would be expected to have the largest hydrodynamic radius.

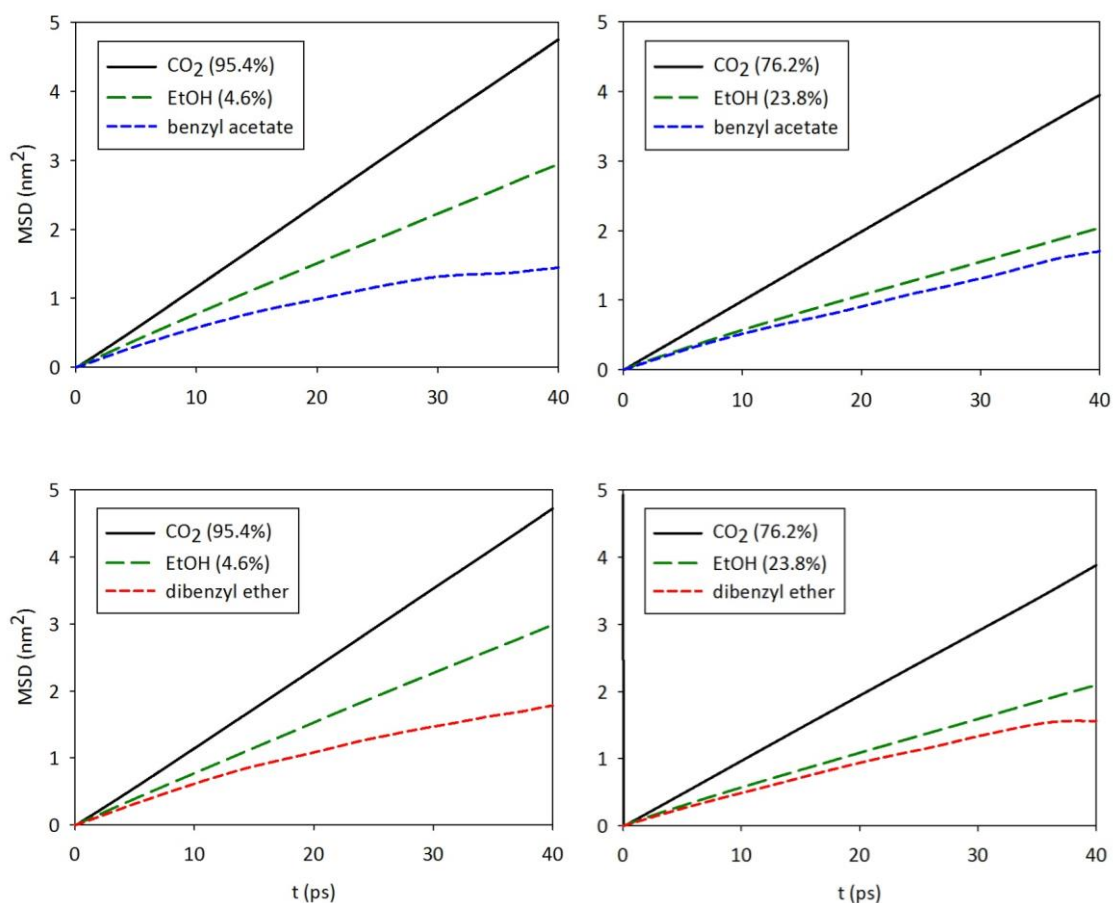


Figure 10.10. Mean squared displacements for the three components in each simulation.

Of greatest interest is the effect of ethanol mole fraction on solute diffusion coefficients. As concentration fraction is increased, the simulated solute diffusion coefficient decreases substantially, indicating slowed solution motion – decreasing by 11% for the polar solute benzyl acetate, but by 15% for dibenzyl ether. Although the magnitude of change of diffusion coefficients as determined from simulation is not substantially different, the experimentally measured differences in diffusion coefficients also differ by roughly the same amount – in that case, 25-27%. In both cases, the diffusion coefficient of the solute begins to approach the diffusion coefficient of ethanol,

suggesting that the stronger ethanol-solute interactions play a significant role in governing solute behavior.

Table 10.8. Diffusion coefficients for all three components within each simulation, with comparison to experimentally determined solute diffusion coefficients from Suarez-Iglesias et al.³⁴

Solute	EtOH mole fraction	Diffusion coefficients ($\times 10^{-9} \text{ m}^2/\text{s}$)			
		CO ₂	EtOH	solute	solute (experimental)
Benzyl acetate	4.8%	19.9	12.5	8.3	8.88 ± 0.16
Benzyl acetate	23.6%	16.6	8.7	7.4	6.60 ± 0.04
Dibenzyl ether	4.8%	19.6	12.7	9.2	8.08 ± 0.20
Dibenzyl ether	23.6%	16.5	8.9	7.8	5.86 ± 0.07

10.7 Discussion

Based on this data, it can be seen that polar ethanol molecules within sc-CO₂/EtOH mixtures preferentially cluster around polar solute molecules. This is particularly evidence around sites where hydrogen bonding is possible, and is not seen around uncharged sites in each solute, such as around phenyl ring carbons. The local mole fraction of ethanol around these sites rises to several times the bulk concentration in every case.

Hydrogen-bonding is noticed around every oxygen site in both solutes simulated, although to the greatest extent around the ether oxygen in dibenzyl ether. However,

because of multiple hydrogen-bonding sites on benzyl acetate, composition differences appear to have less of an impact on first-solvation-shell behavior around benzyl acetate than around dibenzyl ether. Molecules in the first solvation shell of a given solute behave similarly – both in terms of solvation shell dynamics and hydrogen bond dynamics – despite differences in bulk solvent composition, with generally slightly longer lifetimes.

Finally, simulations were able to reasonably reproduce experimentally determined solute diffusion coefficients. With increasing ethanol mole fraction, solute diffusion coefficients were slowed and become more comparable to diffusion coefficients for ethanol. This is consistent with the “entraining” effect proposed to help explain the behavior of solutes in CO₂/polar cosolvent mixtures.

Chapter 11: Further Research Questions

This work has attempted to describe the structure and dynamics of CO₂/ethanol mixtures and their interactions with solute molecules. One of the primary topics we have explored regarding solvation in supercritical systems is the mechanisms governing solvent and solute mobility in a mixture. The slower diffusion of ethanol in sc-CO₂/ethanol mixtures can be explained, as described in Chapter 2, by an increasing number of hydrogen bonds and increasing size of hydrogen-bonded aggregates, each member of which exerts a retarding effect on the motion of its hydrogen-bonded neighbors. Slower diffusion of solute molecules can be similarly explained, as hydrogen bonding increases in prevalence with increasing ethanol mole fraction, retarding solute motion.

Although the behavior of ethanol was studied as a function of the size and structure of larger hydrogen-bonded aggregates, comparable analyses have not been carried out for solutes. This remains a question for further study: do solute molecules which participate in hydrogen bonding exhibit similar behavior when part of heterogeneous hydrogen-bonded networks of molecules? It seems likely that diffusion and rotational motion of solute molecules is affected by this longer-range structure, but this becomes complicated by factors such as the possible presence of multiple hydrogen bonding sites on solutes, or the presence of sites which can only function as Lewis bases within a hydrogen bond (such as the oxygen within dibenzyl ether previously discussed).

This study was also limited to examination of one cosolvent and two different solutes; this obviously suggests pathways for further extension. Given the fact that

ethanol's strongest interactions with solutes through the limited number of hydrogen-bonding sites – and not with the remainder of the solute molecules – it would be worthwhile to examine systems in which both the hydrogen-bonding capacity of the solute and solvent molecules differ. Other polar cosolvents, such as acetone, which cannot function as a proton donor for hydrogen bonding, have been employed in conjunction with supercritical CO₂, and could be simulated and studied in a similar manner to this work; such studies could offer insight into the role of mere attraction between oppositely charged sites on polar molecules (short of full-fledged hydrogen bonding), in solvation within these systems where the cosolvent also appears to act as an entrainer. Similarly, solutes with varying degrees of hydrogen-bonding capacity and varying numbers of possible hydrogen bonds can be simulated and would help to elucidate and expand upon these findings. Finally, in order to generate data that is more statistically reliable, however, simulations of a much greater length than those reported here will be required, in order to effectively sample dynamics of solute molecules in different conditions as the local solvent environment changes.

We have also characterized how changes in thermodynamic conditions (such as composition or pressure) sometimes appear to affect the structural properties of these mixtures without altering the dynamics of molecules with similar local environments. Such behaviors can also be investigated over wider and more varied ranges of conditions, and, in particular, at ranges of values around the critical temperature and pressure, both within the supercritical regime and within the liquid/gas coexistence region, and can help serve to delimit the range of situations in which the conclusions drawn in this study apply.

Contact with experimental data is also crucial for validating theoretical models. Concerted studies in which data is extracted from molecular dynamics simulation and fit to experimental data have successfully accounted for experimental NMR and IR shifts.^{21, 26} However, such studies have not yet attempted to model characteristics of solutions such as hydrogen bond lifetimes, which can be probed using techniques such as 2D-IR spectroscopy.⁵⁵ Furthermore, although solvatochromic experiments with CO₂-expanded liquids have been used to calculate local composition or local density enhancements^{6b, 9, 29, 53}, this has not yet been done in the supercritical regime for multi-component systems. Explicit attempts to simulate systems which have been studied experimentally, experimental work on systems and conditions that have been studied by simulation, or, ideally, the sort of concerted studies describe above, are a necessity to both validate theoretical models and to account for and fully explain experimental observations.

The work described in this dissertation has continued the process of connecting simulation and experimental data for sc-CO₂/cosolvent systems, connecting molecular dynamics trajectories with diffusion data, a physical parameter not previously explored. It also continues to validate the use of computer simulation as a tool for understanding and predicting the complex behavior of mixtures in the supercritical regime, an area that is seeing increasing interest and range of applications.

References: Part II

1. Hannay, J.; Hogarth, J., On the solubility of solids in gases. *Proceedings of the Royal Society of London* **1879**, 30 (200-205), 178-188.
2. Rincon, J.; De Lucas, A.; Garcia, M.; Garcia, A.; Alvarez, A.; Carnicer, A., Preliminary study on the supercritical carbon dioxide extraction of nicotine from tobacco wastes. *Separation Science and Technology* **1998**, 33 (3), 411-423.
3. Hubert, P.; Vitzthum, O. G., Fluid extraction of hops, spices, and tobacco with supercritical gases. *Angew. Chem.-Int. Edit. Engl.* **1978**, 17 (10), 710-715.
4. Zhuze, T. P., Compressed hydrocarbon gases as a solvent. *Petroleum* **1960**, 23, 298-300.
5. Kurnik, R. T.; Reid, R. C., Solubility of solid mixtures in supercritical fluids. *Fluid Phase Equilib.* **1982**, 8 (1), 93-105.
6. (a) Tucker, S. C., Solvent density inhomogeneities in supercritical fluids. *Chem. Rev.* **1999**, 99 (2), 391-418; (b) Kim, S. W.; Johnston, K. P., Clustering in supercritical fluid mixtures. *Aiche J.* **1987**, 33 (10), 1603-1611.
7. Dobbs, J. M.; Wong, J. M.; Johnston, K. P., Nonpolar cosolvents for solubility enhancement in supercritical fluid carbon dioxide. *J. Chem. Eng. Data* **1986**, 31 (3), 303-308.
8. Ke, J.; Mao, C.; Zhong, M. H.; Han, B. X.; Yan, H. K., Solubilities of salicylic acid in supercritical carbon dioxide with ethanol cosolvent. *J. Supercrit. Fluids* **1996**, 9 (2), 82-87.
9. (a) Kelley, S. P.; Lemert, R. M., Solvatochromic characterization of the liquid phase in liquid-supercritical CO₂ mixtures. *Aiche J.* **1996**, 42 (7), 2047-2056; (b)

- Chen, J. W.; Shen, D.; Wu, W. Z.; Han, B. X.; Wang, B.; Sun, D. H., Solvatochromic behavior of phenol blue in CO₂+ethanol and CO₂+n-pentane mixtures in the critical region and local composition enhancement. *J. Chem. Phys.* **2005**, *122* (20), 204508-204514.
10. Ting, S. S. T.; Macnaughton, S. J.; Tomasko, D. L.; Foster, N. R., Solubility of naproxen in supercritical carbon dioxide with and without cosolvents. *Ind. Eng. Chem. Res.* **1993**, *32* (7), 1471-1481.
 11. Jessop, P. G.; Subramaniam, B., Gas-expanded liquids. *Chem. Rev.* **2007**, *107* (6), 2666-2694.
 12. Fulton, J. L.; Yee, G. G.; Smith, R. D., Hydrogen-bonding of methyl alcohol-D in supercritical carbon dioxide and supercritical ethane solutions. *J. Am. Chem. Soc.* **1991**, *113* (22), 8327-8334.
 13. Lalanne, P.; Tassaing, T.; Danten, Y.; Cansell, F.; Tucker, S. C.; Besnard, M., CO₂-ethanol interaction studied by vibrational spectroscopy in supercritical CO₂. *J. Phys. Chem. A* **2004**, *108* (14), 2617-2624.
 14. Kanakubo, M.; Aizawa, T.; Kawakami, T.; Sato, O.; Ikushima, Y.; Hatakeda, K.; Saito, N., Studies on solute-solvent interactions in gaseous and supercritical carbon dioxide by high-pressure ¹H NMR spectroscopy. *J. Phys. Chem. B* **2000**, *104* (12), 2749-2758.
 15. Schnabel, T.; Srivastava, A.; Vrabec, J.; Hasse, H., Hydrogen bonding of methanol in supercritical CO₂: Comparison between ¹H NMR spectroscopic data and molecular simulation results. *J. Phys. Chem. B* **2007**, *111* (33), 9871-9878.

16. Taylor, C. M. V.; Bai, S.; Mayne, C. L.; Grant, D. M., Hydrogen Bonding of Methanol in Supercritical CO₂ Studied by ¹³C Nuclear Spin-Lattice Relaxation. *The Journal of Physical Chemistry B* **1997**, *101* (29), 5652-5658.
17. (a) Curtiss, L. A., Molecular-orbital studies of methanol polymers using a minimal basis set. *J. Chem. Phys.* **1977**, *67* (3), 1144-1149; (b) Hagemester, F. C.; Gruenloh, C. J.; Zwier, T. S., Density functional theory calculations of the structures, binding energies, and infrared spectra of methanol clusters. *J. Phys. Chem. A* **1998**, *102* (1), 82-94; (c) Dixon, J. R.; George, W. O.; Hossain, M. F.; Lewis, R.; Price, J. M., Hydrogen-bonded forms of methanol IR spectra and abinitio calculations. *Journal of the Chemical Society, Faraday Transactions* **1997**, *93* (20), 3611-3618.
18. Zhang, Y.; Yang, J. C.; Yu, Y. X.; Li, Y. G., Structural and hydrogen bond analysis for supercritical ethanol: A molecular simulation study. *J. Supercrit. Fluids* **2005**, *36* (2), 145-153.
19. Stubbs, J. M.; Siepmann, J. I., Binary phase behavior and aggregation-of dilute methanol in supercritical carbon dioxide: A Monte Carlo simulation study. *J. Chem. Phys.* **2004**, *121* (3), 1525-1534.
20. Xu, W. H.; Yang, J. C.; Hu, Y. Y., Microscopic structure and interaction analysis for supercritical carbon dioxide-ethanol mixtures: a Monte Carlo simulation study. *J. Phys. Chem. B* **2009**, *113* (14), 4781-4789.
21. Idrissi, A.; Vyalov, I.; Kiselev, M.; Jedlovszky, P., Assessment of the potential models of acetone/CO₂ and ethanol/CO₂ mixtures by computer simulation and thermodynamic integration in liquid and supercritical states. *Physical Chemistry Chemical Physics* **2011**, *13* (36), 16272-16281.

22. Skarmoutsos, I.; Dellis, D.; Samios, J., Investigation of the local composition enhancement and related dynamics in supercritical CO₂-cosolvent mixtures via computer simulation: The case of ethanol in CO₂. *J. Chem. Phys.* **2007**, *126* (22), 224503-224512.
23. Skarmoutsos, I.; Guardia, E.; Samios, J., Hydrogen bond, electron donor-acceptor dimer, and residence dynamics in supercritical CO₂-ethanol mixtures and the effect of hydrogen bonding on single reorientational and translational dynamics: A molecular dynamics simulation study. *J. Chem. Phys.* **2010**, *133* (1).
24. Chatzis, G.; Samios, J., Binary mixtures of supercritical carbon dioxide with methanol. A molecular dynamics simulation study. *Chem. Phys. Lett.* **2003**, *374* (1-2), 187-193.
25. Aida, T.; Aizawa, T.; Kanakubo, M.; Nanjo, H., Relation between volume expansion and hydrogen bond networks for CO₂-alcohol mixtures at 40 degrees C. *J. Phys. Chem. B* **2010**, *114* (43), 13628-13636.
26. Reiser, S.; McCann, N.; Horsch, M.; Hasse, H., Hydrogen bonding of ethanol in supercritical mixtures with CO₂ by ¹H NMR spectroscopy and molecular simulation. *The Journal of Supercritical Fluids* **2012**, *68*, 94-103.
27. Li, H.; Maroncelli, M., Solvation and solvatochromism in CO₂-expanded liquids. 1. simulations of the solvent systems CO₂+cyclohexane, acetonitrile, and methanol. *The Journal of Physical Chemistry B* **2006**, *110* (42), 21189-21197.
28. Li, H.; Arzhantsev, S.; Maroncelli, M., Solvation and solvatochromism in CO₂-expanded liquids. 2. Experiment-simulation comparisons of preferential solvation in three prototypical mixtures. *The Journal of Physical Chemistry B* **2007**, *111* (12), 3208-3221.

29. Swalina, C.; Arzhantsev, S.; Li, H.; Maroncelli, M., Solvation and solvatochromism in CO₂-expanded liquids. 3. The dynamics of nonspecific preferential solvation. *The Journal of Physical Chemistry B* **2008**, *112* (47), 14959-14970.
30. Gohres, J. L.; Kitchens, C. L.; Hallett, J. P.; Popov, A. V.; Hernandez, R.; Liotta, C. L.; Eckert, C. A., A spectroscopic and computational exploration of the cybotactic region of gas-expanded liquids: methanol and acetone. *The Journal of Physical Chemistry B* **2008**, *112* (15), 4666-4673.
31. Phillips, D. J.; Brennecke, J. F., Spectroscopic measurement of local compositions in binary liquid solvents and comparison to the NRTL equation. *Ind. Eng. Chem. Res.* **1993**, *32* (5), 943-951.
32. (a) Sahena, F.; Zaidul, I. S. M.; Jinap, S.; Karim, A. A.; Abbas, K. A.; Norulaini, N. A. N.; Omar, A. K. M., Application of supercritical CO₂ in lipid extraction - A review. *J. Food Eng.* **2009**, *95* (2), 240-253; (b) Cansell, F.; Aymonier, C., Design of functional nanostructured materials using supercritical fluids. *J. Supercrit. Fluids* **2009**, *47* (3), 508-516; (c) Cooper, A. I., Recent developments in materials synthesis and processing using supercritical CO₂. *Adv. Mater.* **2001**, *13* (14), 1111-1114; (d) Moribe, K.; Tozuka, Y.; Yamamoto, K., Supercritical carbon dioxide processing of active pharmaceutical ingredients for polymorphic control and for complex formation. *Adv. Drug Deliv. Rev.* **2008**, *60* (3), 328-338.
33. Skerget, M. S. M.; Knez, Z.; Knez-Hrncic, M., Solubility of Solids in Sub- and Supercritical Fluids: a Review. *J. Chem. Eng. Data* **2011**, *56* (4), 694-719.
34. Suarez-Iglesias, O.; Medina, I.; Pizarro, C.; Bueno, J. L., Diffusion of benzyl acetate, 2-phenylethyl acetate, 3-phenylpropyl acetate, and dibenzyl ether in

- mixtures of carbon dioxide and ethanol. *Industrial & Engineering Chemistry Research* **2007**, *46* (11), 3810-3819.
35. Yonker, C. R.; Smith, R. D., Effect of the partial molar volume of the solute in the stationary phase on retention in supercritical fluid chromatography. *Journal of Chromatography* **1988**, *459*, 183-191.
 36. Chalaris, M.; Samios, J., Hydrogen bonding in supercritical methanol. A molecular dynamics investigation. *J. Phys. Chem. B* **1999**, *103* (7), 1161-1166.
 37. Smith, W.; Yong, C. W.; Rodger, P. M., DL_POLY: Application to molecular simulation. *Mol. Simul.* **2002**, *28* (5), 385-471.
 38. Allen, M. P.; Tildesley, D.J., *Computer Simulation of Liquids*. Oxford University Press: New York, 1987.
 39. Tian, Y.-l.; Chen, L.; Li, M.-w.; Fu, H.-f., Calculation of gas-liquid critical curves for carbon dioxide-1-alkanol binary systems. *The Journal of Physical Chemistry A* **2003**, *107* (17), 3076-3080.
 40. Harris, J. G.; Yung, K. H., Carbon dioxide's liquid-vapor coexistence curve and critical properties as predicted by a simple molecular model. *J. Phys. Chem.* **1995**, *99* (31), 12021-12024.
 41. (a) Jorgensen, W. L.; Madura, J. D.; Swenson, C. J., Optimized intermolecular potential functions for liquid hydrocarbons. *J. Am. Chem. Soc.* **1984**, *106* (22), 6638-6646; (b) Jorgensen, W. L., Optimized intermolecular potential functions for liquid alcohols. *J. Phys. Chem.* **1986**, *90* (7), 1276-1284.
 42. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *Journal of Computational Physics* **1977**, *23* (3), 327-341.

43. (a) Nosé, S., A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511-519; (b) Hoover, W. G., Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* **1985**, *31* (3), 1695.
44. Melchionna, S.; Ciccotti, G.; Holian, B. L., Hoover NPT dynamics for systems varying in shape and size. *Molecular Physics* **1993**, *78* (3), 533-544.
45. Roney, A. B.; Space, B.; Castner, E. W.; Napoleon, R. L.; Moore, P. B., A molecular dynamics study of aggregation phenomena in aqueous n-propanol. *J. Phys. Chem. B* **2004**, *108* (22), 7389-7401.
46. (a) Chen, B.; Potoff, J. J.; Siepmann, J. I., Monte Carlo calculations for alcohols and their mixtures with alkanes. Transferable potentials for phase equilibria. 5. United-atom description of primary, secondary, and tertiary alcohols. *J. Phys. Chem. B* **2001**, *105* (15), 3093-3104; (b) Bernardi, E.; Stassen, H., Molecular dynamics simulations of acetonitrile/dimethyl sulfoxide liquid mixtures. *J. Chem. Phys.* **2004**, *120* (10), 4860-4867.
47. (a) Ladanyi, B. M.; Skaf, M. S., Computer simulation of hydrogen-bonding liquids. *Annu. Rev. Phys. Chem.* **1993**, *44*, 335-368; (b) Beveridge, D. L.; Mezei, M.; Mehrotra, P. K.; Marchese, F. T.; Ravishanker, G.; Vasu, T.; Swaminathan, S., Monte-Carlo computer simulation studies of the equilibrium properties and structure of liquid water. *Advances in Chemistry Series* **1983**, (204), 297-351.
48. Sum, A. K.; Sandler, S. I., Ab initio calculations of cooperativity effects on clusters of methanol, ethanol, 1-propanol and methanethiol. *J. Phys. Chem. A* **2000**, *104* (6), 1121-1129.

49. Pitman, M. C.; Suits, F.; Gawrisch, K.; Feller, S. E., Molecular dynamics investigation of dynamical properties of phosphatidylethanolamine lipid bilayers. *J. Chem. Phys.* **2005**, *122* (24), 244715-244724.
50. Saiz, L.; Padro, J. A.; Guardia, E., Dynamics and hydrogen bonding in liquid ethanol. *Molecular Physics* **1999**, *97* (7), 897-905.
51. Rapaport, D. C., Hydrogen bonds in water network organization and lifetimes. *Molecular Physics* **1983**, *50* (5), 1151-1162.
52. (a) Foster, N. R.; Singh, H.; Yun, S. L. J.; Tomasko, D. L.; Macnaughton, S. J., Polar and nonpolar cosolvent effects on the solubility of cholesterol in supercritical fluids. *Ind. Eng. Chem. Res.* **1993**, *32* (11), 2849-2853; (b) Li, Q. S.; Zhang, Z. T.; Zhong, C. L.; Liu, Y. C.; Zhou, Q. R., Solubility of solid solutes in supercritical carbon dioxide with and without cosolvents. *Fluid Phase Equilib.* **2003**, *207* (1-2), 183-192.
53. Li, H. P.; Arzhantsev, S.; Maroncelli, M., Solvation and solvatochromism in CO₂-expanded liquids. 2. Experiment-simulation comparisons of preferential solvation in three prototypical mixtures. *J. Phys. Chem. B* **2007**, *111* (12), 3208-3221.
54. (a) Wick, C. D.; Martin, M. G.; Siepmann, J. I., Transferable potentials for phase equilibria. 4. United-atom description of linear and branched alkenes and alkylbenzenes. *J. Phys. Chem. B* **2000**, *104* (33), 8008-8016; (b) Stubbs, J. M.; Potoff, J. J.; Siepmann, J. I., Transferable potentials for phase equilibria. 6. United-atom description for ethers, glycols, ketones, and aldehydes. *J. Phys. Chem. B* **2004**, *108* (45), 17596-17605.

55. Woutersen, S.; Mu, Y.; Stock, G.; Hamm, P., Hydrogen-bond lifetime measured by time-resolved 2D-IR spectroscopy:*N*-methylacetamide in methanol. *Chemical Physics* **2001**, *266* (2), 137-147.

APPENDIX A

FULL STUDENT TEXT OF
“WHAT HAPPENS WHEN CHEMICAL COMPOUNDS
ARE ADDED TO WATER” (WEEK III)
LABORATORY ACTIVITY

What happens when substances are added to water?

Seth Anthony, Nancy J. Boldt, and Dawn Rickey

Department of Chemistry, Colorado State University, Fort Collins, CO 80523-1872

Initial Model (“Elaborate Your Model” from previous week’s experiments)

1. Based on your generalized refined model, draw molecular-level views of what happens when the soluble solids $\text{Pb}(\text{NO}_3)_2$, KI , NaOH , and K_2SO_4 are added (separately) to water.
2. Elaborate your model to include mixtures of different substances that have been added to water – this will function as an “initial model” for next week’s experiments. First, describe your general understanding of what happens when aqueous solutions of different substances are mixed. Then, specifically discuss two mixtures: (1) $\text{Pb}(\text{NO}_3)_2(\text{aq})$ mixed with $\text{KI}(\text{aq})$, and (2) $\text{NaOH}(\text{aq})$ mixed with $\text{K}_2\text{SO}_4(\text{aq})$. Describe your understanding of what would happen if you mixed these solutions. Include both macroscopic- and molecular-level descriptions **before, during, and after** mixing these solutions in your model.

Part VI. What happens when $Pb(NO_3)_2(aq)$ and $KI(aq)$ are mixed?

Observe Label three test tubes (A, B, and C). Using the same balance, *mass each test tube three times* to the nearest 0.01 g **before** adding any of the solutions. Use clean, dry 10-mL graduated cylinders to add the quantities of solutions indicated below to the test tubes. With clean, dry stirring rods, stir each mixture thoroughly, being careful to minimize any loss of material. Record your observations.

Test Tube A: 4.0 mL 0.1 M $Pb(NO_3)_2(aq)$, 2.0 mL 0.1 M $KI(aq)$

Test Tube B: 3.0 mL 0.1 M $Pb(NO_3)_2(aq)$, 3.0 mL 0.1 M $KI(aq)$

Test Tube C: 2.0 mL 0.1 M $Pb(NO_3)_2(aq)$, 4.0 mL 0.1 M $KI(aq)$

Next, centrifuge the contents of the test tubes. (Be sure that all tubes in the centrifuge are balanced before switching it on.) While you are waiting for the centrifuging to finish, obtain some additional 0.1 M $Pb(NO_3)_2(aq)$ and 0.1 M $KI(aq)$ in small vials, and measure and record the conductivities of the solutions. Also record any other observations you make.

Once the centrifuging is complete, use plastic micro-pipettes to siphon off as much of the liquid as possible from each test tube into (separate) clean, labeled vials. Put the test tubes into a beaker and place the beaker into a heated oven, noting the time. Measure the conductivity of each sample of liquid siphoned off.

If time permits, quantitatively determine the mass of precipitates formed by drying them. If you do not have time to dry precipitates, record your qualitative observations about the amount of precipitate formed in each test tube. After your test tubes have been in the oven for about 10 minutes, carefully remove the beaker containing the test tubes from the oven. (*Remember that both the beaker and the test tubes will be very hot, so do not handle with bare hands!*) Observe the contents of your test tubes closely: do you see any water or wet solid present? (Be sure to record all observations in your notebook.) If so, place the test tubes back into the oven for another 5 minutes and repeat the process. Once the contents of your test tubes appear dry, let the test tubes cool, and using the same balance that you used to mass the empty test tubes, *mass each test tube three times to the nearest 0.01 g*. Then put the beaker of test tubes back into the oven. After another 5 minutes, remove the test tubes from the oven, let them cool, and using the same balance, *mass each test tube three times*. If the average mass of a test tube is the same as the average mass of its previous weighing (to within 0.02 g), consider the precipitate to be dry. If the average mass of a test tube is different than it was on previous weighing (by more than 0.02 g), put that test tube back in the oven for another 5 minutes and repeat the process until the mass is constant to within 0.02 g. Be sure to record and clearly label all mass measurements you take during this process in your notebook.

If you have time while waiting for the contents of these test tubes to dry, we recommend that you begin Part VII of the experiment.

Reflect Based on your data and observations, what happens *on the molecular level* when $\text{Pb}(\text{NO}_3)_2(\text{aq})$ and $\text{KI}(\text{aq})$ are mixed? Discuss this with your group. Start by summarizing your data and observations, and comparing your data to the data posted by other groups in your class. What types of evidence did you obtain? What does each piece of evidence tell you about what happened? What trends or patterns were consistent across all groups in your class? Construct a detailed scientific explanation of what happens on the molecular level that is well-supported by the evidence. Propose one or more ways to write a chemical equation that describes what happens. (Which way best describes what happens?)

Did the mixture of $\text{Pb}(\text{NO}_3)_2(\text{aq})$ and $\text{KI}(\text{aq})$ behave as you predicted? Are your experimental results consistent with your initial model? If your results are not consistent with your model, how would you revise your model? Explain your data and observations as best you can.

Explain Participate in a class discussion about what happens when $\text{Pb}(\text{NO}_3)_2(\text{aq})$ and $\text{KI}(\text{aq})$ are mixed.

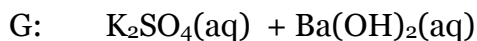
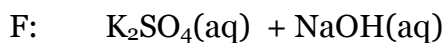
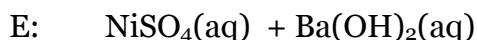
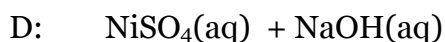
Part VII. Use your refined model to predict how new systems will behave.

Predict

using your

Model Based on your refined model, what do you think will happen when $\text{NiSO}_4(\text{aq})$ and $\text{Ba}(\text{OH})_2(\text{aq})$ are mixed? What do you expect to observe before and after mixing the two solutions? What do you expect to be happening on the molecular level before and after mixing the two solutions? Use pictures to represent your molecular-level view.

Observe Consider what will happen when you mix the pairs of solutions indicated below. In each case, if you assume that a reaction will occur, what ratio of solution volumes would you need to mix for each reactant to be completely consumed?



Measure the conductivities of the four reactant solutions, and use your model to predict what you expect to observe (both visually, and with respect to conductivity) when you perform each reaction.

In four clean, labeled, and massed test tubes, mix the pairs of reactant solutions indicated in D-G above using the ratios you determined, making sure that each test tube will contain a total of 6.0 mL once both reactants are added. Stir and centrifuge the contents of the test tubes. Record all of your observations, including visual observations of relative amounts of any precipitates formed, as well as the conductivities of the resulting solutions. Compare your results to those of other groups to find consistent trends or patterns. If time permits, determine the masses of any precipitates formed.

Reflect Based on your data and observations, what is the identity of each precipitate that formed? What chemical species (molecules, atoms, ions) were present in each solution after any reaction occurred? Discuss your evidence and your reasoning with your group.

Are your experimental results consistent with your revised molecular-level model? If your results are not consistent with your model or if you have things to add, how would you revise your model further? Explain your data and observations as best you can.

Explain Participate in a class discussion about what happens when aqueous solutions are mixed.

Part VIII. What happens when amounts of reactants are varied? (if time permits)

Predict

using your

Model For the reaction between $\text{NiSO}_4(\text{aq})$ and $\text{Ba}(\text{OH})_2(\text{aq})$, what do you expect, both macroscopically and on the molecular level, if you mix varied amounts of each reactant? Choose at least four different ratios of reactants and predict what will occur when different ratios of reactants are mixed.

Observe Follow the same procedure as in Part VII, using the ratios of reactants that you indicated above. Record your macroscopic observations, including conductivity measurements and the masses of products formed.

Reflect What changes when different ratios of reactants are mixed? Are your experimental results consistent with your molecular-level model? If your results are not consistent with your model or if you have things to add, how would you revise your model further? Explain your data and observations as best you can.

Final Refined Model

Develop a final refined model of what happens when aqueous solutions of different chemical compounds are mixed. For (1) $\text{Pb}(\text{NO}_3)_2(\text{aq})$ mixed with $\text{KI}(\text{aq})$, and (2) $\text{NaOH}(\text{aq})$ mixed with $\text{K}_2\text{SO}_4(\text{aq})$ – both before and after each pair of solutions are mixed – present your final macroscopic model (your observations) and your final molecular-level model that accounts for your observations. Compare this model to your previous model (“elaborate your model” from last week), and identify the key aspects of your model that changed and remained the same for both the macroscopic and molecular levels. Fully explain what revisions, if any, you made to your previous model, including what specific experimental evidence has caused you to make any revisions and what has supported your previous refined model in aspects that you did not revise. Based on your experimental evidence, generalize your model so that it could be used to understand what will happen when any aqueous solutions are mixed. Propose a next experiment that would help you further refine or test the molecular-level model you presented above. Explain what the results of this proposed experiment would tell you about what happens on the molecular level when substances are added to water.

APPENDIX B

INSTRUCTOR GUIDE FOR
“WHAT HAPPENS WHEN CHEMICAL COMPOUNDS
ARE ADDED TO WATER” (WEEK III)
LABORATORY ACTIVITY

With Elaborate Your Model:

Discussion of Post-Lab Questions from Previous Week:

- Ask students to share their descriptions of the systems discussed in post-lab “elaborate your model” question #2. Be sure to have students describe (and draw on the board) what they expect to happen before, during, and after each pair of solutions are mixed.
- Common ideas students exhibit include: formation of precipitates (some students may be familiar with solubility rules from past experience, others may make inferences about solubility based on the previous week’s activities); all ionic species remaining in solution (some students may omit spectator ions from their descriptions); reactants undergoing metathesis and forming new “molecular” species that remain in solution.
- During the discussion, ask students to describe what they expect to observe macroscopically (particularly visual observations and conductivity measurements) for each molecular-level prediction. As usual, do not validate any particular answer, but ask students to consider what evidence would allow them to distinguish between different molecular-level models.

With Part VI:

Before beginning the experiment, discuss with students the new experimental techniques that will be involved, including centrifuging and drying precipitates. Drying precipitates is not strictly required during this lab, as students can get a sense of which reaction mixtures produce the most and least precipitate visually. Based on whether your labs tend to run shorter or longer than others, decide whether to have your students dry compounds yourself. Another option is to have a group who you know works quickly and efficiently dry the compounds and report the masses to the class. It will be particularly important for students to read the directions fully before proceeding to avoid missing crucial steps (such as massing test tubes before adding the solutions).

With Observe (“If time permits...”)

A precipitate is considered “dry” when its mass stabilizes. Once the mass stabilizes (stops decreasing), that is the final mass. While precipitates are drying, have each group present their conductivity data on the board/overhead and discuss the trends and patterns observed. Have groups predict which test tubes will contain the highest and lowest masses of precipitate. Even if the drying is not performed, students should still be able to judge relative amounts of precipitates visually.

Have each group share their data on the board so that trends and patterns can be clearly identified. Students should observe the lowest conductivity and greatest precipitate mass in test tube C, the stoichiometric reaction mixture. The non-stoichiometric mixtures with excess reagents result in greater numbers of ions remaining in solution.

Encourage students to consider limiting and excess reagents and how that relates to the mass of precipitate formed and the amount of conductivity. Also encourage students to write chemical equations describing what they believe occurs in this reaction and to draw molecular-level pictures of each mixture, paying close attention to the quantity of each species present.

The precipitate is lead iodide, PbI_2 , but determining the precipitate’s identity is not a central goal of this activity. If students have not yet learned solubility rules, they may be able to identify the precipitate formed based on the mass of precipitate. PbI_2 has a much larger formula weight than KNO_3 ; thus by calculating theoretical yields, only one precipitate becomes reasonable.

With Part VII, Predict:

Students often seek out help with equation writing; however, be aware that writing equations for “double displacement”/”metathesis” reactions is a skill many students are conversant at from high school chemistry, and students can often help each other with this.

With Observe, (“In four clean, labeled, and massed...”)

Students can use balanced chemical equations or molecular-level pictures (paying close attention to quantities and ratios of reactants) to determine the stoichiometric ratios of reactants. Have student groups present their ratios of solution volumes prior to experimentation and discuss reasons for any differences in ratios presented. (This should lead to consideration of using balanced chemical equations for those that did not.)

Test Tube	Volume Ratio Needed	Precipitate Formed	Theoretical Yield of Precipitate
D	2.0 mL : 4.0 mL	Ni(OH) ₂	0.0037 g
E	3.0 mL : 3.0 mL	Ni(OH) ₂ and BaSO ₄	0.019 g
F	N/A (2.0 mL : 4.0 mL from stoichiometry)	no precipitate	no precipitate
G	3.0 mL : 3.0 mL	BaSO ₄	0.014 g

Students are also asked to predict the conductivities of product solutions after measuring the conductivity of reactants. One “reasonable guess” might be that, if no precipitate is formed, the conductivity of the product solution should be roughly halfway between that of the two reactant solutions. Students should see this generally validated for reaction mixture F; other reaction mixtures should produce conductivities well below this theoretical value, indicating to students that something different happens in these mixtures (consistent with the visual appearance of the precipitate).

With Explain:

As before, have students compile their data on the board/overhead in order to identify and discuss trends and patterns observed.

Note that the actual yield of the hydroxides might be greater than the theoretical yields. This may be due to the formation of hydrates. The trends in the data, however, should match expectations from the expected precipitate and the mole ratios in the balanced equations.

As in Part VI, students may be able to determine the identity of the precipitates without knowing the solubility rules. For example, nickel solids/solutions will be blue/green; the mass of precipitates will be the most for the system with 2 precipitates (E); the conductivity should be the lowest for the system with 2 precipitates (E) in which all ions have (ideally) precipitated out of solution.

Since drying sometimes takes longer than expected, students may not always have time to complete these additional experiments during the lab period. Note how long drying solutions has taken students in the previous two activities before deciding whether to proceed with this portion of the experiment.

If some time remains, but not enough to fully dry the precipitates formed in part VIII, have students only test their conductivities of the reaction mixtures.

With Part VIII:

As students experiment with different ratios, they should observe that precipitate mass is lower and conductivities are markedly higher for non-stoichiometric mixtures, with these differences becoming greater the further students are from the stoichiometric 1:1 ratio.

Refined model assignment: point out to students that, instead of focusing on salt and sugar, their refined model assignment asks them to consider the two reactions that they discussed in the “elaborate your model” portion of the previous assignment.

With Final Refined Model Assignment:

I suggest pointing out, regarding the “elaborate” portion, that this model has actually provided identifies of chemical species. Additionally, since different reaction stoichiometries are to be discussed, students may choose to discuss one reaction (such as “A”) in detail, and then identify how they expect the others to differ.

APPENDIX C

FULL STUDENT TEXT OF
“LEARNING TO CONSTRUCT AND REFINE MODELS”
INTRODUCTORY ACTIVITY

Learning to Construct and Refine Models

Using the Model-Observe-Reflect-Explain (MORE) Thinking Frame

Seth Anthony, Laura Wally, and Dawn Rickey

Department of Chemistry, Colorado State University, Fort Collins, CO 80523-1872

This semester, using the Model-Observe-Reflect-Explain (MORE) Thinking Frame, you will construct initial and refined models for the chemical systems you study in the laboratory. The quality of these models will be evaluated as the major part of your laboratory grade. Today's activities provide opportunities for you to learn how to write quality models and to understand the criteria upon which they will be evaluated.

Activity I – Evaluating Example Initial Models

The primary objective of writing an initial model is to describe your current understanding, from both macroscopic and molecular-level perspectives, of the system to be investigated. The ideas that you express in your initial models will not be judged based on whether or not they are scientifically correct. Instead, initial models are evaluated based on their completeness.

Here is an example of an initial model assignment. Read it carefully.

Describe your understanding of what happens when equal amounts of ethanol and water are mixed. Describe what you expect to observe (see, hear, feel, smell) before and after the two liquids are mixed; this is your *initial macroscopic model*. Then, explain what you think the molecules, atoms, and/or ions are doing that results in your observations; this is your *initial molecular-level model*. Describe how you think your molecular-level model is connected to or explains your macroscopic model. In addition to words, you may choose to use pictures to represent your models. (Sometimes a picture is worth a thousand words.) This can be especially helpful for expressing your molecular-level ideas. Molecules, atoms, and ions can be represented with symbols. If you use symbols other than chemical formulas, be sure to include a key.

Your lab manual packet contains some example initial models written in response to the assignment above for you to read and evaluate using the grading rubric provided. Form groups of 3 students and discuss your evaluations of each initial model with your group. Try to reach a consensus as to how you would grade each aspect of each model. Afterwards, participate in a discussion with your entire class about the important aspects of initial models.

Activity II – Evaluating Example Refined Models

Here is a corresponding example of a refined model assignment. Read it carefully.

Develop a refined model for what happens when ethanol and water are mixed, including the following aspects. You may find it easier to integrate these aspects when you are constructing your refined model and we encourage you to do so.

Present your refined model.

For the mixing of equal amounts of ethanol and water, briefly describe your refined macroscopic model (your observations) and your refined molecular-level model that accounts for your observations. Your refined model must be consistent with all of your experimental evidence (observations, data, calculations, and results).

Explain why your model has (or has not) changed from initial to refined.

Compare your refined model with your initial model, and identify key aspects of your model that changed or remained the same. Be specific. Then explain why your model has changed in these ways (or why it has not, if aspects of your refined model are exactly the same as in your initial model) by using specific experimental evidence to support or refute the claims you made in your initial model. (Using evidence in this way is an important part of a good scientific explanation!)

Generalize your refined model.

Based on your experimental evidence, generalize your model so that the model could be used to understand what will happen when any substances (not just ethanol and

water) are mixed. Use specific evidence (from your experiment and/or everyday life) to support your general model.

Propose a next experiment

What questions do you have about what happens when substances are mixed?

Propose a next experiment that would help you further refine or test the molecular-level model you presented above. Explain what the results of this proposed experiment would tell you about what happens on the molecular level when substances are mixed.

After viewing the demonstration that your instructor performs, read and evaluate example refined using the grading rubric provided. Discuss your evaluations of each model with your group. Try to reach a consensus as to how you would grade each aspect of each model. Afterwards, participate in a discussion with your entire class about the important aspects of refined models.

Activity III – Constructing Your Own Initial Model

Construct Your Initial Model of a Burning Candle

Instructions:

Describe your understanding of what happens when a candle burns. Describe what you expect to observe (see, hear, feel, smell) before and after the candle wick has been lit; this is your *initial macroscopic model*. Then, explain what you think the molecules, atoms, and/or ions are doing that results in your observations; this is your *initial molecular-level model*. Describe how you think your molecular-level model is connected to or explains your macroscopic model. In addition to words, you may choose to use pictures to represent your models. (Sometimes a picture is worth a thousand words.) This can be especially helpful for expressing your molecular-level ideas. Molecules, atoms, and ions can be represented with symbols. If you use symbols other than chemical formulas, be sure to include a key.

Activity IV - Evaluating Peers' Initial Models

Instructions:

1. Exchange models with other members of your group without removing it from your notebook (so that peer comments will also be copied). All members of the group should read all models.
2. As you read each of your peers' models, write feedback on their papers about what you think could be improved in each model and about what you think they did well. Be sure to use a different-colored pen and write your name below your feedback.
3. In your group, discuss each initial model. Identify the macroscopic and molecular-level aspects that each student presents, and consider the completeness of each aspect. Does each model connect macroscopic observations with molecular-level behavior? Explain and justify your decisions.
4. Evaluate each of the models in your group using the criteria outlined in the grading rubric.
5. Discuss with members of your group how well each model meets the criteria outlined in the grading rubric. Develop a consensus within your group as to how you would grade each model.
6. Your instructor will then lead a class discussion, with some student ideas displayed on the board.

Activity V – Making Observations and Reflecting

Conduct the following experiment in the laboratory. Both during and following your experiment, reflect upon what is happening. Are your observations consistent with your initial model? Does your initial model fully explain your observations? What do you think is happening on the molecular level?

1. Obtain matches and a candle attached to the bottom of a pie pan. Light the candle. What do you notice? Record your observations.
2. Add water, about $\frac{1}{4}$ of an inch deep, to the pie pan containing the candle. Invert a 250-mL Erlenmeyer flask over the lit candle, making sure that the Erlenmeyer flask's opening is also submerged under the water. Make careful observations and record them.
3. Make sure that all matches are extinguished before they are discarded into the waste baskets. Save the candles for reuse.

Activity VI – Constructing Your Own Refined Model

Develop a refined model for what happens when a candle burns. Be sure to incorporate the following four aspects into your model. ***The first three aspects should be included in this and all refined models that you construct throughout the semester. The “propose a next experiment” aspect is required in the final refined model for each laboratory module.*** You may find it easier to integrate these aspects when you are constructing your refined model and we encourage you to do so. (We have presented the aspects separately below to make it clear what should be included in a refined model.)

Present your refined model.

For a candle before and after the wick is lit, briefly describe your refined macroscopic model (your observations) and your refined molecular-level model that accounts for your observations. Your refined model must be consistent with all of your experimental evidence (observations, data, calculations, and results).

Explain why your model has (or has not) changed from initial to refined.

Compare your refined model with your initial model, and identify the key aspects of your model that changed or remained the same. Be specific. Then explain why your model has changed in these ways (or why it has not, if aspects of your refined model are exactly the same as in your initial model) by using specific experimental evidence to support or refute the claims you made in your initial model. (Using evidence in this way is an important part of a good scientific explanation!)

Generalize your refined model.

Based on your experimental evidence, generalize your model so that the model could be used to understand what will happen when any substance (not just a candle) burns. Use specific evidence (from your experiment and/or everyday life) to support your general model.

Propose a next experiment

What questions do you have about what happens when substances burn? Propose a next experiment that would help you further refine or test the molecular-level model you presented above. Explain what the results of this proposed experiment would tell you about what happens on the molecular level when substances burn.

Activity VII – Evaluating Peers’ Refined Models

In your groups of 3 students, share and discuss your refined models.

1. In your group, discuss what you found to be the most challenging aspects of writing refined models.
2. Exchange models with other members of your group without removing it from your notebook (so that peer comments will also be copied). As you read each of your peers’ models, write feedback on their papers about what you think could be improved in each model and about what you think they did well. Be sure to use a different-colored pen and write your name below your feedback.
3. Then discuss the refined models with other group members. For each refined model, indicate whether or not each of the four important aspects (presenting, explaining, generalizing, and proposing) are present, as well as the quality of each of the aspects.
4. Using the grading rubric provided by your instructor, evaluate each of the models in your group using the criteria outlined in the grading rubric.
5. Discuss with members of your group how well each model meets the criteria outlined in the grading rubric. Develop a consensus within your group as to how you would grade each model. Discuss how well the rubric matches your ideas from your previous discussion about whether a model is complete.
6. Your instructor will then lead a class discussion, and students’ macroscopic and molecular ideas, as well as the changes that have been made to initial models, will be displayed on the board.

7. Based on the class and group discussions and feedback you have just received, make a list of what changes you could make to improve your refined model.

****BEFORE YOU LEAVE LAB TODAY, YOU MUST TURN IN YOUR INTIAL MODEL AND YOUR REFINED MODEL****

Activity VIII – Homework Assignment to Improve Your Refined Model

Further

Now that you've engaged in discussions and received specific feedback about what constitutes a complete, high-quality refined model, you will revise your model of what happens when substances burn that you wrote during class to improve it further. (This revision should be based on critiques and feedback you receive about the quality of your model. Additional experiments are not required.)

To receive full credit for this assignment, you must visit your teaching assistant (or another TA for this course if you are not available during your TA's office hours) in the departmental TA room to discuss aspects of your refined model before revising it. Be sure that the TA signs your laboratory notebook page at the end of this visit.

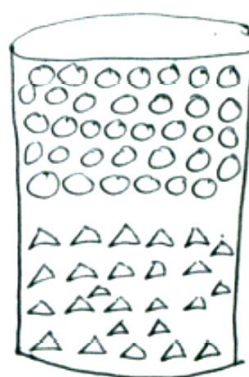
This assignment is due at the beginning of your next laboratory session. At this time, you should submit your second refined model that has been revised based on the feedback you received from your peers and your TA. Be sure to include the page that has been signed by your TA.

Sample Initial Model 1:

Initial Model:

When you add water and ethanol, you will see both liquids in one container, but one will be on top of the other. The ethanol will probably be on bottom and the water will be on top, but it could be the other way. Between the two, you will see a line at the boundary, because they don't mix,

Molecularly, there is no interaction between the water and ethanol. The two do not mix with each other or react. The water molecules repel the ethanol molecules and vice versa, which is why they stay separate and why you can't see that the two liquids don't mix.



○ = water
△ = ethanol

Sample Initial Model 2:

Initial Model

Macroscopic: I expect that water will be clear and I expect that ethanol will be colored. I expect to hear nothing from both liquids. I expect that I will not smell the water, but that the ethanol will smell sweet. I do not expect a taste because we are not supposed to taste chemicals in lab, but I do know that water would have no taste. If I was to touch either, it would feel wet, but the ethanol would be more slippery.

If you add water to ethanol, you will see, hear, and sense all the same things, but there will be twice as much liquid.

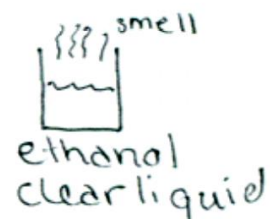
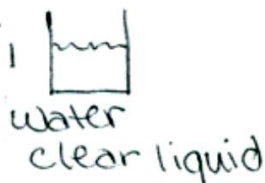
Molecular: When you have water and ethanol separate, both are made up of molecules that easily move past each other. This ability to move around is what makes them liquid. When you mix the two, you will still have molecules moving past each other in the same way as before, and so the observations will be the same as before mixing - you will have a liquid.

Sample Initial Model 3:

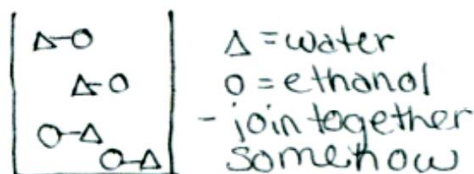
Initial Model:

I believe that before mixing ethanol and water, you will see two clear liquids. The water will not have a smell and the ethanol will smell strong. Neither will make any sounds. The molecules of water are all moving around past each other.

Before:



When the two liquids are mixed, you will see a clear liquid with twice as much volume as the original liquids. It will not look different, but it will have a smell that is half as strong as the original ethanol. They may make a splashing sound as you mix them. The molecules of water and ethanol will react with each other and join together to form a new compound that has properties which are somewhere between the two original compounds.



Sample Model 1:

Reviewed By: _____

Initial Model Grading Rubric

Present your initial model

Describes macroscopic model	Not present	Present but incomplete	Present and complete
Describes molecular-level model	Not present	Present but incomplete	Present and complete

Sample Model 2:

Initial Model Grading Rubric

Present your initial model

Describes macroscopic model	Not present	Present but incomplete	Present and complete
Describes molecular-level model	Not present	Present but incomplete	Present and complete

Sample Model 3:

Initial Model Grading Rubric

Present your initial model

Describes macroscopic model	Not present	Present but incomplete	Present and complete
Describes molecular-level model	Not present	Present but incomplete	Present and complete

Sample Refined Model 1:

Refined Model:

When you add water and ethanol, before mixing them you will see both liquids as clear, odorless, and colorless. They will flow like liquids and be slippery to the touch. After you add water to ethanol, you will see a clear odorless liquid, similar to before. Ethanol and water appear to mix and that take less space than there was before.

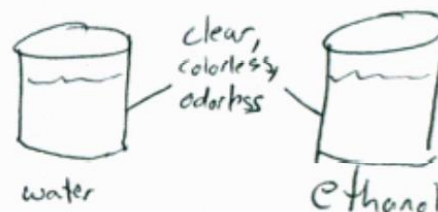
The ethanol molecules probably pair up with the water molecules in close bonds, but do not react. Instead, the water interacts with the ethanol and they are attracted to each other.

I was wrong when I said that the ethanol and water would not mix. I changed my model by adding more detail.

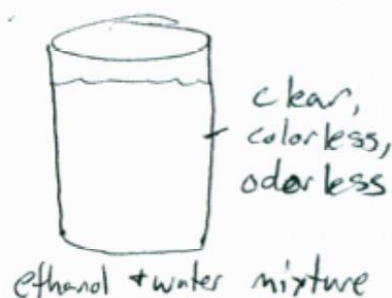
I also was wrong ~~at~~ when I said that the water would not interact with the ethanol. This is different from what I said before, and I changed my model because I saw volume decreases from 50mL to 48mL.

Macroscopic-Level

Before Mixing



After Mixing



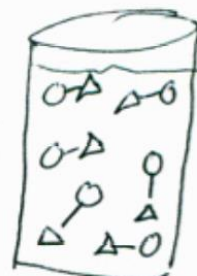
Molecular-Level

Before Mixing



After Mixing

water = O
ethanol = Δ



✗ In general, I believe that, when two liquids are mixed together, they will either pair up with each other or they do not mix and bond. If they bond, then the volume will ~~be~~ be less than before, but if they do not mix, then the volume will be the same. If they don't mix, they'll form two separate layers like when I have seen oil and water poured together.

I am curious to find out if there are other kinds of compounds that when mixed pair up and bond. I would like to test organic liquids, like different kinds of oils. If they mix without a visible line between them but don't compress together and take up less space, then that would indicate that they don't pair up, according to my model.

Sample Refined Model 2:

Refined Model

Macroscopic: Before water and ethanol are mixed, both will be clear liquids without smells or sounds. After they are mixed, the mixture appears the same - clear, with no smell. The mixture takes up less volume than twice the sum of the original volumes.

Molecular: Before mixing, water and ethanol are both liquids and made up of separate molecules moving past each other randomly. When water and ethanol are mixed, most of the water and ethanol molecules move freely around each other, causing the mixture to be a liquid. After mixing, some of the water molecules must be closer to the ethanol molecules than before. They may be bonded or they may just be attracted to each other, but I am uncertain.

Changes to macroscopic model: My initial model was correct, about my observations of smell - both compounds were odorless before and after being mixed - but not about visual observations. Ethanol was a clear liquid and is not colored like I had said initially.

Changes to molecular model: In my initial model, I said that, both before and after being mixed, the water and ethanol molecules just moved past each other. I now know that there must be some form of interaction or attraction, because of what I saw in lab.

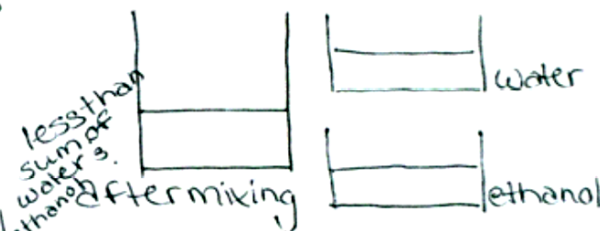
Generalize: It is difficult to predict what will happen when two liquids are mixed, but I expect that the volume of the mixture will depend on what kind of interaction there is between the two things mixed. The stronger the interaction, the closer they will come together, and the smaller the total volume. We saw this in the experiment as well.

New experiment: I'm curious what might happen with other alcohols, such as if we add equal amounts of methanol or butanol to water. Would they have stronger or weaker reactions? We could determine this by measuring the volume of the mixture.

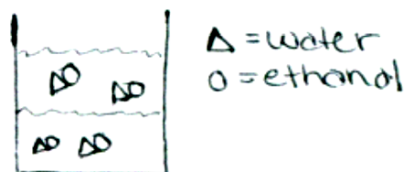
Sample Refined Model 3:

Refined Model:

Unlike in my previous model, I see now that water mixed with ethanol will result in a mixture that takes up less volume than the sum of the two original liquids. However, I was correct about my prediction that there would still be a strong smell after they are mixed.



On the molecular level, my model has not changed. I still believe that compounds ~~there~~ when mixed, join together to form a new compound that has a mixture of the two properties. Because the water has joined together with the ethanol to form one new compound, there is now less space between the water and ethanol portions. I understand now that this is why the volume is less than before. I also believe that my model is supported because I saw wavy lines in the water where the two were joining together and reacting



Refined model continued:
Generally, when compounds are mixed with water, they will react and join together. The visual and other properties will be the average of the two originals. I believe this because I observed the smell of the mixture to be mild, between the smells of ethanol and water and because the appearance of the mixture was the same as both ethanol and water. I also know that when colored compounds are added to water, the water becomes slightly colored, showing that the properties of the mixture fall in between the properties of the things being mixed.

I wonder if this is true for all compounds, especially solids. If you add enough of a solid, will its properties average with the liquid you add it to? To test this, I would mix a strong smelling solid with water, to see if the volume is less than the sum, and to see if the smell varies in proportion to the amount of solid added. This would tell me if everything reacts and joins together on the molecular level in the same way when mixed.

Sample Refined Model 1

Reviewed By: _____

Present your refined model

Describes macroscopic model	Not present	Present but incomplete	Present and complete
Describes molecular-level model	Not present	Present but incomplete	Present and complete
Molecular-level model is consistent with experimental evidence	Not consistent		Consistent

Explain why your model has changed

Identifies key changes to macroscopic model	Not present	Present but incomplete	Present and complete
Identifies key changes to molecular-level model	Not present	Present but incomplete	Present and complete
Uses specific experimental evidence to explain changes to molecular-level model	Not present	Present but incomplete	Present and complete

Generalize your model

Provides generalized model that can be used to predict new situations.	Not present	Present but incomplete	Present and complete
Uses specific experimental evidence to support generalized model	Not present	Present but incomplete	Present and complete

Propose a next experiment

Proposes a specific new experiment	Not present	Present but incomplete	Present and complete
Explains how results of proposed experiment would inform molecular-level understanding of system	Not present	Present but incomplete	Present and complete

Sample Refined Model 2

Reviewed By: _____

Present your refined model

Describes macroscopic model	Not present	Present but incomplete	Present and complete
Describes molecular-level model	Not present	Present but incomplete	Present and complete
Molecular-level model is consistent with experimental evidence	Not consistent		Consistent

Explain why your model has changed

Identifies key changes to macroscopic model	Not present	Present but incomplete	Present and complete
Identifies key changes to molecular-level model	Not present	Present but incomplete	Present and complete
Uses specific experimental evidence to explain changes to molecular-level model	Not present	Present but incomplete	Present and complete

Generalize your model

Provides generalized model that can be used to predict new situations.	Not present	Present but incomplete	Present and complete
Uses specific experimental evidence to support generalized model	Not present	Present but incomplete	Present and complete

Propose a next experiment

Proposes a specific new experiment	Not present	Present but incomplete	Present and complete
Explains how results of proposed experiment would inform molecular-level understanding of system	Not present	Present but incomplete	Present and complete

Sample Refined Model 3

Reviewed By: _____

Present your refined model

Describes macroscopic model	Not present	Present but incomplete	Present and complete
Describes molecular-level model	Not present	Present but incomplete	Present and complete
Molecular-level model is consistent with experimental evidence	Not consistent		Consistent

Explain why your model has changed

Identifies key changes to macroscopic model	Not present	Present but incomplete	Present and complete
Identifies key changes to molecular-level model	Not present	Present but incomplete	Present and complete
Uses specific experimental evidence to explain changes to molecular-level model	Not present	Present but incomplete	Present and complete

Generalize your model

Provides generalized model that can be used to predict new situations.	Not present	Present but incomplete	Present and complete
Uses specific experimental evidence to support generalized model	Not present	Present but incomplete	Present and complete

Propose a next experiment

Proposes a specific new experiment	Not present	Present but incomplete	Present and complete
Explains how results of proposed experiment would inform molecular-level understanding of system	Not present	Present but incomplete	Present and complete

Student Name: _____

Reviewed By: _____

Initial Model Grading Rubric

Present your initial model

Describes macroscopic model

Not
present

Present but
incomplete

Present
and complete

Describes molecular-level model

Not
present

Present but
incomplete

Present
and complete

Student Name: _____

Initial Model Grading Rubric

Present your initial model

Describes macroscopic model

Not
present

Present but
incomplete

Present
and complete

Describes molecular-level model

Not
present

Present but
incomplete

Present
and complete

Student Name: _____

Initial Model Grading Rubric

Present your initial model

Describes macroscopic model

Not
present

Present but
incomplete

Present
and complete

Describes molecular-level model

Not
present

Present but
incomplete

Present
and complete

Student Name: _____ Reviewed By: _____

Present your refined model

Describes macroscopic model	Not present	Present but incomplete	Present and complete
Describes molecular-level model	Not present	Present but incomplete	Present and complete
Molecular-level model is consistent with experimental evidence	Not consistent		Consistent

Explain why your model has changed

Identifies key changes to macroscopic model	Not present	Present but incomplete	Present and complete
Identifies key changes to molecular-level model	Not present	Present but incomplete	Present and complete
Uses specific experimental evidence to explain changes to molecular-level model	Not present	Present but incomplete	Present and complete

Generalize your model

Provides generalized model that can be used to predict new situations.	Not present	Present but incomplete	Present and complete
Uses specific experimental evidence to support generalized model	Not present	Present but incomplete	Present and complete

Propose a next experiment

Proposes a specific new experiment	Not present	Present but incomplete	Present and complete
Explains how results of proposed experiment would inform molecular-level understanding of system	Not present	Present but incomplete	Present and complete

Student Name: _____ Reviewed By: _____

Present your refined model

Describes macroscopic model	Not present	Present but incomplete	Present and complete
Describes molecular-level model	Not present	Present but incomplete	Present and complete
Molecular-level model is consistent with experimental evidence	Not consistent		Consistent

Explain why your model has changed

Identifies key changes to macroscopic model	Not present	Present but incomplete	Present and complete
Identifies key changes to molecular-level model	Not present	Present but incomplete	Present and complete
Uses specific experimental evidence to explain changes to molecular-level model	Not present	Present but incomplete	Present and complete

Generalize your model

Provides generalized model that can be used to predict new situations.	Not present	Present but incomplete	Present and complete
Uses specific experimental evidence to support generalized model	Not present	Present but incomplete	Present and complete

Propose a next experiment

Proposes a specific new experiment	Not present	Present but incomplete	Present and complete
Explains how results of proposed experiment would inform molecular-level understanding of system	Not present	Present but incomplete	Present and complete

Student Name: _____ Reviewed By: _____

Present your refined model

Describes macroscopic model	Not present	Present but incomplete	Present and complete
Describes molecular-level model	Not present	Present but incomplete	Present and complete
Molecular-level model is consistent with experimental evidence	Not consistent		Consistent

Explain why your model has changed

Identifies key changes to macroscopic model	Not present	Present but incomplete	Present and complete
Identifies key changes to molecular-level model	Not present	Present but incomplete	Present and complete
Uses specific experimental evidence to explain changes to molecular-level model	Not present	Present but incomplete	Present and complete

Generalize your model

Provides generalized model that can be used to predict new situations.	Not present	Present but incomplete	Present and complete
Uses specific experimental evidence to support generalized model	Not present	Present but incomplete	Present and complete

Propose a next experiment

Proposes a specific new experiment	Not present	Present but incomplete	Present and complete
Explains how results of proposed experiment would inform molecular-level understanding of system	Not present	Present but incomplete	Present and complete

APPENDIX D

INSTRUCTOR GUIDE FOR
“LEARNING TO CONSTRUCT AND REFINE MODELS”
INTRODUCTORY ACTIVITY

Instructors' Guide: Learning to Construct and Refine Models

Using the Model-Observe-Reflect-Explain (MORE) Thinking Frame

Seth Anthony, Laura Wally, and Dawn Rickey

Department of Chemistry, Colorado State University, Fort Collins, CO 80523-1872

Major learning goals for this activity:

- What is the molecular level; how one can communicate about it clearly and concisely?
- What is MORE; what will be expected as part of assignments (particularly initial and refined models) in this course? How does this enact the “scientific method”?

Other items of interest to discuss with students:

- Instructor's expectations about how students will structure their lab notebook pages.
- Instructor's expectations about how students will write and organize models.
- Molecular level ideas about the mixing of liquids.
- Molecular level ideas about candle burning and combustion.

Activity I – Evaluating Example Initial Models (Duration: ~30 minutes)

Summary: Students individually review the initial model assignment and three sample initial models provided for a chemical system (the mixing of equal volumes of ethanol and water), rating the quality of various aspects of the model on a version of the grading rubric used during the semester. Students then discuss their ratings in small groups,

attempting to reach a consensus on their evaluations as preparation for a whole-class discussion.

Instructor Notes:

- All activities (except for activity V) should be conducted in an area appropriate for small group discussion if possible, rather than at laboratory benches.
- After reviewing sample models individually, students should work in groups of 3 to discuss their evaluations of them; form groups of 4 only when necessary.
- Instruct students to try to reach a consensus with fellow group members about how they evaluated the models; if some groups reach a consensus more quickly than others, they can debate with neighboring groups to reach a multi-group consensus. Encourage students to identify and try to agree on aspects that they think are molecular. Allow about 5-10 minutes for students to discuss their evaluation of the sample models.
- For class discussion, talk about each aspect of each model explicitly, asking students to cite portions of the model to justify their grading choice. Ask whether aspects are present or not and have students support their evaluations with examples; ask students to contrast models. Possible questions to use during class discussion include:
 - How did you grade this aspect? Was it present or not?
 - Why did you think it was complete or incomplete?
 - How could the student have improved this aspect of the model?

- Which of the sample models exemplifies this aspect the best?
- When students seek standards for how models will be graded, attempt to have them reach a consensus about how they would grade something themselves before providing normative answers; and emphasize ways in which even “complete” model aspects can be improved. Students will often be rather “strict” in grading these sample models; this can be encouraged as it is likely to help set a higher bar for students' own models.

Notes on sample initial models:

Sample Initial Model 1

The macroscopic model is present but incomplete. It is incomplete because it does not address ethanol and water before they are mixed as requested in the assignment.

(It is very common for model assignments to ask students to consider things before, during, and/or after an event, so it will be helpful to call students attention to this.)

In addition, the student only considers visual observations. This is a less serious omission, but we want to encourage students to think about the variety of observations that they might make. *(It is also common that students do not expect to observe anything for a particular sense or senses. If that is the case, then they can write things like, “I do not expect to hear or smell anything before ethanol and water are mixed.”)*

The molecular-level model is also present but incomplete because it does not address ethanol and water before they are mixed. There are also some things that would improve the molecular-level model after mixing. For example, in the key for the picture, if the student had indicated that circle equals “a water molecule” and triangle equals “an ethanol molecule,” it would have been clearer that the student’s picture is intended as a molecular-level representation (“water” and “ethanol” could refer to the macro level). In general, we want students to be talking about specific particles when they discuss the molecular level. The sentence “The water molecules repel the ethanol molecules...” is a good example of a clear molecular-level description in words.

Sample Initial Model 2

The macroscopic model is present and complete.

The initial molecular-level model would also be considered to be present and complete, but there are still some things that would improve it. It would be helpful to include a molecular-level picture as this could provide more detail about how the student thinks the molecules might be arranged. The molecular-level model could also be more detailed in other ways. For example, does the student have ideas about what the water and ethanol molecules are composed of?

Sample Initial Model 3

The macroscopic model is present and complete.

The molecular-level model is present and complete. More detail about the molecular-level view of ethanol and water before mixing, including a picture, would improve the model, as would a clearer key identifying water and ethanol molecules within the molecular-level picture. Also, there are no explicit connections made between the macroscopic and molecular-level models, so that is an aspect that could be improved. This is not something we take points off for in initial models, but we do want to encourage students to try to make connections between the macroscopic and molecular levels.

Activity I – Evaluating Example Initial Models (continued)

With all students wearing safety glasses, perform the ethanol/water mixing demo before moving on to critiquing refined models. Pour 25 mL of ethanol into the 50 mL graduated cylinder and 25 mL of water in the 25 mL graduated cylinder; show students the volumes and discuss how to measure quantities in glassware by reading from the bottom of the meniscus. Then pour the water into the ethanol. Volumes must be precisely measured if you are to clearly observe the desired outcome – a non-additive (less than 50 mL) volume of the resulting mixture. Take the mixture after you have used it for the demonstration to the lab and empty it into liquid waste to model proper waste disposal for students.

Activity II – Evaluating Example Refined Models (Duration: ~40 minutes) -

Summary: As with initial models, students individually review a refined model assignment and three sample refined models, rating the quality of various aspects using the grading rubric, and then discuss their ratings – first within small groups, then as a class.

Instructor Notes:

- Activity II proceeds similarly to Activity I above. Useful questions include:
 - How did you grade this aspect? Was it present or not?
 - Why did you think it was complete or incomplete?
 - How could the student have improved this aspect of the model?
 - Which of the sample models exemplifies this aspect the best?

- Allow slightly longer for students to review the refined models and to discuss their evaluations than they needed for their initial model reviews.
- Again, emphasize the molecular-level and the use of language explicitly referring to molecules, atoms, ions, etc., and how pictures can help make molecular-level ideas clear.
- Emphasize discussion of model changes and supporting evidence and how this parallels the practices of scientists in revising their theories and explanations in response to observations. Evidence plays a key role throughout the refined model assignment: it is what students describe within the macroscopic portion of their models, what supports changes discussed in the refined model (or aspects that don't change), what supports generalization, and what should be the focus of discovery and interpretation in their proposed new experiment.
- One common student question that emerges during this activity: How can changes be discussed if a model doesn't change? Use this question as an opportunity to discuss how evidence can either lead to refinement or support previously-stated aspects of a model – or both! In either case, whether an aspect of an initial model has been preserved, has changed, or is entirely new, that can be identified and then should be supported with evidence.

Notes on sample refined models:

Sample Refined Model 1

The macroscopic model is present and complete.

The molecular-level model is present and complete and consistent with the experimental evidence. It would be improved by specifying molecules in the picture keys and providing more detail.

Key changes to the macroscopic model are present but incomplete since the student does not specify what details were added to the macroscopic model, such as a description of the two compounds before mixing, or the observations related to volume. Also, in the statement, “I changed my model by adding more detail,” it is not explicit that the student is referring to the macroscopic model; this has to be inferred by the reader. To improve, the student could have said something like “I also added descriptions of ethanol and water before mixing as well as information about the volume change after mixing to my macroscopic model.” (Note that students do not need to indicate every change, but should be specific about important changes.)

Key changes to the molecular-level model is present and complete. However, in the last paragraph/picture in the section, this student never clearly indicates that he/she is talking about molecules. (It can be inferred from the second paragraph of the refined model in conjunction with the pictures.) So an important improvement here would be to specify that: “I also was wrong when I said that the water molecules would not interact with the ethanol molecules.”

Uses specific evidence to explain changes is present but incomplete. Although specific evidence is cited to justify the change, the student does not fully explain why

that evidence prompted the particular change in the molecular-level model. A high-quality scientific explanation does not simply claim that evidence supports a particular model; it lays out the relationship between the cited evidence and the model it is claimed to support. *(In grading, this example would be at the high end of the present but incomplete category since the student does link specific evidence to the model change, but does not fully explain why this evidence prompted the model change.)* To earn a “present and complete,” the student would need to say something like: "I changed my molecular-level model because I saw the volume decrease from 50 mL to 48 mL when ethanol and water were mixed. *The fact that the liquids occupied less space when they are mixed indicates some sort of attraction between the water and ethanol that pulled the molecules closer together than when the liquids were separate.* Thus, I think that the two liquids definitely interacted, unlike what I said in my initial model."

Provides generalized model is present and complete. This aspect would be improved if the student provided guidelines for predicting under what circumstances two liquids would mix or not mix, but this information would likely be difficult for general chemistry students to glean from the single example used for the demo. In the laboratory modules they will carry out, they will typically have more information to discern such patterns. *(Note also that this student uses relevant everyday evidence in addition to experimental evidence. This is not only permissible, but encouraged.)*

Uses specific evidence to support generalized model is present and complete.

Sample Refined Model 1 (continued)

Proposes specific experiment is present and complete. More details, such as indicating the specific pairs of substances that would be mixed, would improve it.

Explains how experiment would explain molecular model is present and complete. More clarity and specificity about the molecular-level would improve it. (e.g., “...that would indicate that they don’t pair up...” is not clearly describing molecular-level behavior; this has to be inferred from previous model text. Writing “...that would indicate that molecules of substance 1 do not pair up with molecules of substance 2...” would add clarity.)

Sample Refined Model 2

The macroscopic model is present and complete.

The molecular-level model is present and complete and consistent with the experimental evidence. As in the corresponding initial model, it would be improved by including a picture.

Key changes to the macroscopic model are present but incomplete since the student does not address whether the key experimental observation regarding volume changed or remained the same. Also, the student is not accurate about the smell observations since he/she indicated that ethanol would smell sweet in the initial model. Accurate reflection on changes would therefore improve the model. (*Note: The accuracy of the reported changes is not something that TAs are required to grade, since it would be necessary to look back at the initial models to assess this. However, it*

is very important for students' learning that they accurately reflect on model changes, so instructors should emphasize to them that they should read their previous models before writing about changes to their models to help insure that they are accurate.)

Key changes to the molecular-level model are present and complete.

Uses specific evidence to explain changes is not present. (*Although it is common for students to make vague references to evidence like "because of what I saw in lab," students must cite specific evidence in order for this aspect to be considered present.*)

Provides generalized model is present and complete. This aspect would be improved if the student provided guidelines for predicting under what circumstances there would be stronger vs. weaker interactions, but this information would likely be difficult for general chemistry students to glean from the single example used for the demo. (In the laboratory modules they will carry out, they will typically have more information to discern such patterns.)

Uses specific evidence to support generalized model is present but incomplete (or perhaps not present). Since there was only one experiment in this case, we know which experiment the student must be referring to, but the student does not describe specific evidence.

Proposes specific experiment is present and complete.

Explains how experiment would explain molecular model is not present.

Sample Refined Model 3

The macroscopic model is present but incomplete since it does not address ethanol and water before they are mixed.

The molecular-level model is present and complete and consistent with the evidence.

Key changes to the macroscopic model are present and complete.

Key changes to the molecular-level model are present and complete.

Uses specific evidence to explain changes is present and complete. (*Note that in this case the student uses specific evidence to explain why her molecular-level model has not changed.*)

This is a good example of a model that integrates the presentation of the refined model with the explanation of changes (or not) to the model.

Provides generalized model is present but incomplete because the student only addresses substances mixed with water, while the assignment asks students to address any substances being mixed.

Uses specific evidence to support generalized model is present and complete. (*Note that this student uses relevant everyday evidence in addition to experimental evidence. This is not only permissible, but encouraged.*)

Proposes specific experiment is present and complete.

Explains how experiment would explain molecular model is present and complete.

Activity III – Constructing Your Own Initial Model (Duration: ~10 minutes) –

Summary: Responding to an initial model “prompt” much like those they will encounter throughout the semester, students write their own initial model of a new system – a candle both before and after it is lit.

Instructor Notes:

- It is likely that some students will not finish with writing their models in the time allotted. As time runs out, suggest to students who are still working on their models that they put down important points in bullet form, rather than in full sentences and paragraphs.
- Students often ask about appropriate model length; an initial model should be long enough to cover all of the aspects asked for, but this can usually be done in half a page to a page. Stress that content and completeness are more important than length.
- Encourage students to use multiple modes of representing their ideas – words, pictures, and/or chemical equations – as they saw within the sample models they evaluated.

Activity IV - Evaluating Peers' Initial Models (~10 minutes) –

Summary: Within small groups, students exchange the initial models they have just written, evaluating each others' work using a version of the grading rubric, leaving written feedback for peers, and discussing each others' models based on those evaluations.

Instructor Notes:

- Encourage students to “swap” models in their groups as they finish writing them, rather than waiting for the entire class. You may make new groups of students who are done writing the initial model.
- Allow 5-10 minutes for students to evaluate their peers' models.
- Have students use the rubric page 21 to evaluate peers' models; collect this at the end of this activity or at the end of the class period as part of assessing students' class participation.
- If time permits, hold a class discussion in which students volunteer ideas and expectations that they included in their initial models – this can be a good way to get students used to the idea of the initial model discussions that will be held at the beginning of labs, and a way to accustom students to the idea that different students can articulate different initial ideas and still have complete models.

Activity V – Making Observations and Reflecting (Duration: ~25 minutes) –

Summary: In the laboratory, students observe a candle, both before and after it is lit, in order to refine their model and practice recording and reflection on experimental observations. Students also perform an additional experiment in which a flask is inverted over a candle while the candle sits in a pan of water, in order to provide additional evidence about the chemical processes happening during the reaction.

Instructor Notes:

- Discuss lab safety (safety glasses, precautions around flame) before letting the students perform the experiment. Recommend that students rinse their used matches under water prior to throwing them in the trash.
- Since students will be working in groups of four, it is likely that they will need to rearrange groups again. This can be a good opportunity to meet other students in the class.
- Discuss what types of observations students might make, using all appropriate senses (sound, smell, etc.) Encourage students to record all observations, from the “obvious” to the unexpected. Suggest that they compare their observations with other groups in order to learn about things they might have overlooked.
- Ask students explicitly to take note of whether their observations agree with their initial model – is there evidence that they didn't describe or that violates their expectations?
- Prompt students to pay particular attention to evidence that could help them explain what is happening at the molecular level. Encourage them to think about and discuss this as they describe their observations.

- With the flask-and-water experiment, students may need to adjust the water height in the pan depending on the geometry of the pan and candle.
- Students should clean up thoroughly before proceeding to Activity VI. Inspect each student group's lab station and sign students' notebooks when you have approved them to leave the lab.

Activity VI – Constructing Your Own Refined Model (Duration: ~15 minutes) –

Summary: Responding to a refined model “prompt” like those they will continue to encounter, students write their refined models of the burning candle system.

Instructor Notes:

- Students are often eager to charge ahead with model writing after performing the experiment.
- Each student should write their refined model individually, even though they may have discussed changes and evidence as a group. This is likely to take longer than writing their initial model. (~15 minutes)
- Students frequently ask whether they must “re-state” their initial model in their refined model. It’s useful both to refer students to examples from the sample models and to remind them that they now (and will always) have new data and observations to incorporate into their refined models.

Activity VII – Evaluating Peers’ Refined Models (Duration: ~15 minutes) –

Summary: As with initial models, students exchange the refined models they have just written, evaluate each work, leave written feedback, and discuss each others’ models based on those evaluations, prior to a class discussion about their models and brainstorming of possible model improvements.

Instructor Notes:

- Refined model reviewing proceeds similarly to initial model reviewing.
- Students should use the grading rubrics on pages 22-24 to evaluate peers’ models. After the instructor has verified that students have participated in this part of the activity, students can keep the pages that contain the reviews of their in-class models and use this information to inform their revisions in the post-lab assignment.
- Use the feedback portion of the activity to talk about the feedback that will be written on student lab reports, and emphasize that it is important that students read your comments (as opposed to just looking at grades/points) in order to improve their models in the future.
- If time permits, have students who finish first put their models (or aspects of them) on the board. Have them walk the class through their models. Emphasize that they should discuss:
 - Aspects of their model that stayed the same.
 - Aspects of their model that changes.
 - Evidence that supports their discussion of changes

- Remind students that their homework assignment is to rewrite their revised models based on what they've learned. Their list of possible improvements will help them in this assignment.

Activity VIII – Homework Assignment to Improve Your Refined Model

Further

Summary: Based on the feedback and discussion they have had about their refined models during the class period, students' homework assignment is to revise and improve their models of the candle system.

Instructor Notes:

- As you collect student work the end of the lab period, collect the copy of lab notebook pages that do not have peers' notes on them – these notes are useful for students when they revise their refined model. Also remind them that they can visit their instructors' or another instructors' office hours to discuss their model.

APPENDIX E

STUDENT TEXT FOR SECOND WEEK OF
“LEARNING TO CONSTRUCT AND REFINE MODELS”
INTRODUCTORY ACTIVITY
WITH INSTRUCTOR GUIDE TEXT

Instructor guide text is placed in boxes between student text.

Activity I – “What Makes a Good Model?” (~25 minutes)

1. Begin by sharing your initial ideas about what makes a good model with your group. (The initial model assignment is repeated below.)

Imagine that it's early next semester, and you receive an email from a friend who is taking CHEM 112:

“Hey – I’m writing my refined model for this week’s chem lab, but I don’t know if it has everything it’s supposed to have. Can you tell me what things you tried to make sure to do in your models? I know what’s in the assignment, but last time our TA gave us an extra tip: “Try to use multiple senses to describe your macroscopic observations.” Are there other things you always made sure you checked for or questions that you asked yourself when you wrote your refined models? Thanks!”

You don’t know which refined model your friend is working on, so you want to give her advice that’s general, but that will still help her figure out whether or not her model is complete. For your initial model assignment, compose a reply to your friend’s email. In your reply, describe your general understanding of what makes a high-quality refined model of a chemical system. Include at least four

specific suggestions or questions that your friend could ask about her own model to check that her refined model is complete.

2. Discuss the ideas that the individual members of your group came up with for their initial models. In comparing initial models, which ideas are similar to each other and which are distinct? Are there other important criteria for good models that are not represented in individual students' initial models? Make sure you record all of your group's ideas in your lab notebook.
3. Now integrate your group's ideas with your own and make notes on how you would change your e-mail message based on this discussion.
4. After your group has discussed the reply you would send, your TA will lead a class discussion about your ideas, and may ask you to write some of your suggestions on the board. After the discussion, leave your suggestions on the board so you and other groups can refer to them through the entire class period.

Activity II – Evaluating Peers’ Refined Models (~20 minutes)

1. Open your lab notebook to your refined model for “Spectroscopic Analysis of Food Dye, Part 1”. (This is the refined model you turned in last week, based on the experiments completed two weeks ago.)
2. Exchange notebooks with the other members of your group. As you read each of their refined models for “Spectroscopic Analysis of Food Dye, Part 1,” imagine that you are the instructor grading it.
3. Use the grading rubrics on the following pages to evaluate each aspect of the models. For each aspect, leave a comment of one or two sentences suggesting improvements.
4. When you are finished evaluating your group members’ models, keep the completed grading rubrics (don’t show them to anyone yet), and return the notebooks to their owners.

Activity III – Evaluating Your Own Refined Model (~10 minutes)

1. After the other two members of your group have evaluated your model, retrieve your lab notebook.
2. Using the grading rubric provided, evaluate your own model as you evaluated the others. Don’t forget to leave yourself suggestions about how to improve the aspects you rated “present, but incomplete,” or “not present.” Based on what you’ve seen of other students’ work and what you’ve discussed, also indicate how you could make aspects you rate as “present and complete” even better.

Activity IV – Discussing Your Evaluations (~10 minutes)

1. After you have finished evaluating your own model, ask your group members to share their evaluations of your model with you. Compare their ratings and suggestions for improvement with yours. Mark any areas of disagreement.
2. Discuss the ratings in your small group. Discuss what criteria you used to determine if particular aspects were complete or not. When you disagreed on a rating, was it because you were using different criteria to judge that aspect of the model, or for another reason?

Students should discuss their evaluations and suggestions for improvement, additionally have students give the rubric that evaluates the model to the evaluated student so they can have it to refer to as they revise their model.

Reflection Questions

Answer the following reflection questions in your lab notebook.

- Which aspect of writing a refined model do you find the most difficult? Explain your reasoning.
- What ideas did you get from reading other people's models? For instance, what did you learn from observing the way they structured their models or the things that they chose to include that might be useful to think about when you write your own refined models?

Activity V – Examining Instructor Feedback (~15 minutes)

Your instructor will now hand back the graded copies of the “Spectroscopic Analysis of Food Dyes, Part 1” refined models, complete with written feedback.

1. Read over your instructor’s comments and compare them with the comments you wrote to yourself. After you do so, answer the following reflection questions in your notebook:
 - In what categories do your self-ratings most disagree with the grade you earned from your instructor? What do you think caused these differences?
 - Identify any instructor feedback comments that you don’t understand, as well as aspects of the model for which you don't understand why you did not receive full credit. What additional information would help you understand how to improve your models?
2. Participate in a class discussion with your instructor about the feedback they provided and how to determine if your model is complete. This is an opportunity for you to ask questions about aspects of refined models that may be unclear to you.

Students should discuss their evaluations and suggestions for improvement, additionally have students give the rubric that evaluates the model to the evaluated student so they can have it to refer to as they revise their model.

Activity VI – Constructing Evidence-Based, Scientific Models

(Candle Observations, ~30 minutes)

1. In your lab notebook, turn to the last refined model of the candle system that you handed in two weeks ago. Read it over before proceeding. You will be provided with the same laboratory setup as during the original candle experiment.
2. Conduct the following experiments in the laboratory. **Wear safety glasses (of course!) and be sure to keep a safe distance away from all flames.** Both during and after your experiments, reflect upon what you think is happening. Are your observations consistent with your previous refined model? Does your previous model fully explain your observations? What do you think is happening on the molecular level – that is, what are atoms, molecules, and/or ions, doing that explain your observations?
 - Add water, about $\frac{1}{4}$ of an inch deep, to a pie pan with a candle in it.
 - Holding a match about 2 inches from the candle wick, light the match and slowly move the match's flame to the candle wick while carefully observing what happens.
 - Once the candle is lit, hold the flask provided upside-down above the candle flame for about one minute. Then, quickly place the flask down into the water as far away from the candle as possible (not over the candle) so that the opening of the flask is submerged under the water. Make careful observations and record them. Then remove the flask from the water.

Students should see a greater water level rise when the flask is held over the candle for a minute before placed down over it. Encourage students to consider this observation in light of their ideas from their previous initial model – is it consistent with the idea that the water level rises because of the consumption of oxygen?

- For this step, we suggest that two group members work together: While the candle is still burning, have one group member light a second match, while a second group member gently blows the candle flame (but not the match flame) out with a puff of breath. Try to minimize the disturbance to the air surrounding the candle when blowing out the candle. Then take the lit match, and starting by holding it about two inches away from the extinguished candle, move it slowly towards the wick. Try moving the match towards the candle from different directions (for example, through the candle smoke, in the direction opposite the smoke, etc.) Again, make careful observations and record them.

Passing the match through the smoke from a recently extinguished candle should cause the flame to “jump” from the match to the candle and re-ignite the candle. If students do not initially see this, try suggesting that they move the match towards the candle from different directions or at different times – immediately vs. later – after extinguishing the flame. Do not “give away” the role of smoke – this is something students should discover.

- You may conduct additional experiments (from the first week of candle experiments, from this week, or of your own design) to help you understand what is happening if time allows.
- Make sure that all matches are extinguished before they are discarded into the waste baskets. Save the candles for reuse.

Groups that finish this activity earlier than others can begin revising their candle models, either in the lab room or back in the recitation room.

When you go back to the recitation room following this activity, be sure to give the stockroom a “heads up” that you’re out of the lab room (but also do your final checkout with them after your students are gone at the end of the period).

3. Read your previous refined model of the candle system once again.
4. In your lab notebook, write a new “molecular-level” section for your model according to the instructions below:

Present your refined model.

For a candle before and after the wick is lit, briefly describe your refined molecular-level model that accounts for your observations. Your refined model must be consistent with all of your experimental evidence (observations, data, calculations, and results).

Explain why your model has (or has not) changed compared with your previous refined model. Compare your refined molecular-level model presented above with your previous refined model, and identify the key aspects of your model that changed or remained the same. Be specific. Then explain why your model has changed in these ways (or why it has not for key aspects of your refined model that are the same as in your previous refined model) by using specific experimental evidence to support or refute the claims you made in your initial model. (Using evidence in this way is an important part of a good scientific explanation!)

5. After writing this revised molecular-level section for your model of the candle, read the models of the other two members of your group, and then answer the following reflection questions in your laboratory notebook:
 - Which of the models you read presents the most convincing case about what happened on the molecular-level? What aspects of the refined model assignment contribute to making a convincing case?
 - How do you think the aspects of making a convincing case in your refined model relate to what you know about the “scientific method” and how actual science is done?

Activity VII – What Makes a Good Model?" Revisited (~15 minutes)

1. Now refine your initial model by revising the email reply your group wrote at the beginning of the lab period. Having completed today's activities:
 - What suggestions would you add to the list you previously developed?
 - What suggestions would you place more or less importance on than you did before?
 - What suggestions (if any) would you remove from the list?
2. Be sure that each group member records the revised email reply in their notebooks.
3. Finally, participate in a class discussion about how your ideas about writing refined models have changed because of today's activities. Revise the list that your class placed on the board earlier by adding new suggestions that you think are important and drawing a line through (not erasing) items you think aren't as important.

**Activity VIII – Homework Assignment to Revise Your Refined Model from
“Spectroscopic Analysis of Food Dyes, Part 1”**

Now that you’ve engaged in discussions and received specific feedback about what constitutes a complete, high-quality refined model, you will revise your first refined model from the “Spectroscopic Analysis of Food Dyes” module (the model you evaluated and received back graded today) to improve it further. This revision should be based on critiques and feedback you receive about the quality of your model, and the ideas you've thought about today in class. No additional experiments are required.

This assignment is due at the beginning of your next laboratory session.

APPENDIX F

TEXT OF ALL INITIAL AND REFINED MODEL
ASSIGNMENTS FROM FALL 2008 IMPLEMENTATION

Spectroscopy

Barbara A. Reisner

Department of Chemistry, James Madison University, Harrisonburg, VA 22807

Joycette Santos-Santori and Dawn Rickey

Department of Chemistry, Colorado State University, Fort Collins, CO 80523

Initial Model Assignment:

Consider the following scenario in constructing your initial model: An electrical discharge tube is made of a glass tube that contains metal electrodes at each end and is filled with hydrogen (H_2) gas at low pressure. A high voltage is placed across the electrodes, and when the current is switched on, a stream of fast-moving electrons shoots through the gas from the cathode to the anode. This causes the hydrogen (H_2) molecules inside of the tube to break apart into individual H atoms and the electrons in the H atoms to be excited to higher energy levels.

For your initial model, describe your understanding of what will happen next. First, describe what you expect to observe (see, hear, feel, smell) after the current is switched on; this is your *initial macroscopic model*. Then, explain what you think the atoms, molecules, ions, and/or subatomic particles (protons, neutrons, electrons) are doing that results in your observations; this is your *initial molecular-level model*. Describe how you think your molecular-level model is connected to or explains your macroscopic

model. In addition to words, you may choose to use pictures to represent your models. (Sometimes a picture is worth a thousand words.) This can be especially helpful for expressing your molecular-level ideas. Molecular-level species can be represented with symbols. If you use symbols other than chemical formulas, be sure to include a key.

Refined Model Assignment:

Develop a refined model for what happens when a collection of atoms in a discharge tube is excited by high voltage, including the following aspects. You may find it easier to integrate these aspects when you are constructing your refined model and we encourage you to do so.

Present your refined model.

For a discharge tube containing hydrogen, briefly describe your refined macroscopic model (your observations) and your refined molecular-level model that accounts for your observations. Your refined model must be consistent with all of your experimental evidence (observations, data, calculations, and results).

Explain why your model has (or has not) changed from initial to refined.

Compare your refined model with your initial model, and identify the key aspects of your model that changed or remained the same. Be specific. Then explain why your model has changed in these ways (or why it has not, if aspects of your refined model are exactly the same as in your initial model) by using specific experimental evidence to support or refute the claims you made in your initial model. (Using evidence in this way is a very important part of a good scientific explanation!)

Generalize your refined model.

Based on your experimental evidence, generalize your model so that the model could be used to understand what will happen when any collection of atoms in a discharge tube is excited by high voltage. Use specific experimental evidence to support your general model.

The Chemistry of Smell Module

Barbara A. Reisner

Department of Chemistry, James Madison University, Harrisonburg, VA 22807

Dawn Rickey

Department of Chemistry, Colorado State University, Fort Collins, CO 80523

Lydia T. Tien

Department of Chemistry & Geosciences, Monroe Comm. College, Rochester, NY 14623

Angelica M. Stacy

Department of Chemistry, University of California, Berkely, CA 94720

Initial Model Assignment:

Describe your understanding of how humans detect odors and why substances smell the way that they do. What do you think your nose is detecting when you smell something? What makes it possible for you to distinguish different odors? Why might two substances smell different from or similar to each other? Be sure to address both the macroscopic and molecular levels. Describe how you think your molecular-level model is connected to or explains your macroscopic model.

Refined Model Assignment:

Develop a refined model for why substances smell the way they do. The following four aspects should be included in this and all refined models that you construct throughout the semester. You may find it easier to integrate these aspects when you are constructing your refined model and we encourage you to do so. (We have presented the aspects separately below to make it clear what should be included in a refined model.)

Present your refined model.

Briefly describe your refined macroscopic model (your observations) and your refined molecular-level model for why substances smell the way they do that accounts for your observations. Your refined model must be consistent with all of your experimental evidence (observations, data, calculations, and results).

Explain why your model has (or has not) changed from initial to refined.

Compare your refined model with your initial model, and identify the key aspects of your model that changed or remained the same. Be specific. Then explain why your model has changed in these ways (or why it has not, if aspects of your refined model are exactly the same as in your initial model) by using specific experimental evidence to support or refute the claims you made in your initial model. (Using evidence in this way is an important part of a good scientific explanation!)

Generalize your refined model.

Based on your experimental evidence, generalize your model so that the model could be used to understand why any substance (not only the things you have smelled today) smells the way it does. What criteria can you use to predict how something (anything) will smell? Use specific experimental evidence to support your general model.

What happens when substances are added to water?

Adam C. Mattox and Barbara A. Reisner

Department of Chemistry, James Madison University, Harrisonburg, VA 22807

Seth Anthony, Nancy J. Boldt, and Dawn Rickey

Department of Chemistry, Colorado State University, Fort Collins, CO 80523

Initial Model Assignment:

Describe your understanding of what happens to substances when they are added to water. For your initial model, discuss salt (NaCl) and sugar (C₁₂H₂₂O₁₁), two solids that can be found in any kitchen cupboard. Describe what you expect to observe (see, hear, feel, smell) before and after you have added salt to water and before and after you have added sugar to water; this is your *initial macroscopic model*. Then, explain what you think the molecules, atoms, and/or ions are doing that results in your observations; this is your *initial molecular-level model*. Describe how you think your molecular-level model is connected to or explains your macroscopic model.

When developing your model of what happens when substances are added to water, be sure to incorporate responses to the following questions/statements into your model (Your instructor will be looking for these.)

- What kinds of changes do you expect to observe when you put each substance in water?
- How will the amount of substance added affect your observations?

- Using the language of chemistry, explain what is happening to your compounds as they are added to water.
- Draw pictures of what you expect the atoms, ions, and/or molecules to look like before and after the substances have been added to water. Molecules, atoms, and ions can be represented with symbols. If you use symbol other than chemical formulas, be sure to include a key.

Refined Model Assignment:

Develop a refined model for what happens to substances when they are added to water, including the following aspects. You may find it easier to integrate these aspects when you are constructing your refined model and we encourage you to do so.

Present your refined model.

For salt (NaCl) and sugar (C₁₂H₂₂O₁₁), briefly describe your refined macroscopic model (your observations) and your refined molecular-level model that accounts for your observations. Your refined model must be consistent with all of your experimental evidence (observations, data, calculations, and results).

Explain why your model has (or has not) changed from initial to refined.

Compare your refined model with your initial model, and identify the key aspects of your model that changed or remained the same. Be specific. Then explain why your model has changed in these ways (or why it has not, if aspects of your refined model are exactly the same as in your initial model) by using specific experimental evidence to support or refute the claims you made in your initial model. (Using evidence in this way is a very important part of a good scientific explanation!)

Generalize your refined model.

Based on your experimental evidence, generalize your model so that the model could be used to understand what will happen when any substance (not just NaCl or $C_{12}H_{22}O_{11}$) is added to water. What criteria can you use to determine what will happen when different substances are added to water? How does amount affect what happens? Use specific experimental evidence to support your general model.

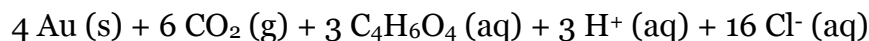
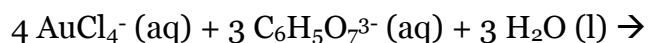
Exploring Gold Nanoparticles

Colin Blair, Ellen R. Fisher, and Dawn Rickey

Department of Chemistry, Colorado State University, Fort Collins, CO 80523

Initial Model Assignment:

During the next laboratory module, you will synthesize gold nanoparticles, Au(s), and explore the factors that affect their properties. In the first week of the module, you will synthesize gold nanoparticles by mixing aqueous solutions of chloroauric acid (HAuCl₄) and sodium citrate (Na₃C₆H₅O₇). The net ionic equation is:



succinic acid

Elaborate your model of what happens when aqueous solutions of different chemical compounds are mixed by describing your understanding of what will happen when you mix HAuCl₄ (aq) and Na₃C₆H₅O₇ (aq) at an elevated temperature. For the macroscopic aspect, describe what you expect to observe for the reactant solutions, during the mixing of the two solutions, and for the resulting product mixture. For the molecular-level aspect, explain what you think the particles (e.g. molecules, atoms, ions) are doing that results in your expected observations before, during, and after mixing these solutions.

Refined Model Assignment:

Develop a refined model of what happens when you mix $\text{HAuCl}_4(\text{aq})$ and $\text{Na}_3\text{C}_6\text{H}_5\text{O}_7(\text{aq})$ at an elevated temperature. Briefly present your refined macroscopic model (your observations) and your refined molecular-level model that accounts for your observations. Be sure to address your understanding of the nature of gold nanoparticles from both macroscopic- and molecular-level perspectives in your model. Compare this model to your initial model, and identify the key aspects of your model that changed or remained the same. Fully explain what revisions, if any, you made to your initial model, including what specific experimental evidence has caused you to make any revisions and what has supported your initial model in aspects that you did not revise. Also explain how each aspect of your refined model (whether the same or different from your initial model) is consistent with your data and observations. Based on your experimental evidence, generalize your model so that it could be used to better understand similar chemical systems.

Chemical Change and Energy:

What Fuel Makes the Best Energy Source?

Barbara A. Reisner, Stuart A. McGee, and A. Colin Mattox

Department of Chemistry, James Madison University, Harrisonburg, VA 22807

Melonie A. Teichert, Jennifer Davis, and Dawn Rickey

Department of Chemistry, Colorado State University, Fort Collins, CO 80523

Initial Model Assignment:

Describe your understanding of how chemical reactions transfer heat. For your initial model, discuss the two reactions that you will study in the first week of this laboratory module: (1) the reaction of NaOH (aq) with HCl (aq) and (2) the reaction of ethanol, $\text{CH}_3\text{CH}_2\text{OH}(\text{l})$ with oxygen gas. For each of these reactions, include macroscopic and molecular-level perspectives before, during, and after each reaction. Describe how you think your molecular-level model is connected to or explains your macroscopic model. Be sure to include information about what you think is responsible for any heat transfer during the reaction from a molecular-level perspective. As you develop this initial model, consider how your models from previous experiments relate to your model of chemical reactions and heat transfer.

Refined Model Assignment:

Referring back to your initial model, develop and briefly present a refined model that accounts for all of your data and observations from today's experiments. Be sure to relate the data and observations to what you think is happening at the molecular level that results in heat transfer. Compare your refined model with your initial model, and identify the key aspects of your model that changed or remained the same. Fully explain what revisions, if any, you made to your initial model, including what specific experimental evidence has caused you to make any revisions and what has supported your initial model in aspects that you did not revise. Also explain how each aspect of your refined model (whether the same or different from your initial model) is consistent with your data and observations. Based on your experimental evidence, generalize your model so that it could be used to understand what happens when any chemical reaction transfers heat.

APPENDIX G

SAMPLE REVIEWING PACKET FROM InSPRE
(IN-CLASS SELF- AND PEER-REVIEWING EXERCISE)
IN-CLASS ACTIVITY

Thermo Model Reviewing – Marie Curie

**Hand your refined model to Svante Arrhenius.
Obtain Harold Urey’s refined model and review it below.**

Review of Harold Urey’s refined model:

	Not present		Present but incomplete		Present and complete
Present your refined model					
Describes macroscopic model	(1)	(2)	(3)	(4)	(5)
Describes molecular-level model	(1)	(2)	(3)	(4)	(5)
Molecular-level model is consistent with experimental evidence	(No)				(Yes)
Explain why your model has changed					
Identifies key changes to macroscopic model	(1)	(2)	(3)	(4)	(5)
Identifies key changes to molecular-level model	(1)	(2)	(3)	(4)	(5)
Uses specific experimental evidence to explain changes to molecular-level model	(1)	(2)	(3)	(4)	(5)
Generalize your model					
Provides generalized model that can be used to predict new situations	(1)	(2)	(3)	(4)	(5)
Uses specific experimental evidence to support generalized model	(1)	(2)	(3)	(4)	(5)
		Not at all accurate		Somewhat accurate	
How accurate do you believe your ratings above are?	(1)	(2)	(3)	(4)	(5)

**Now pass Harold Urey’s model to Ernest Rutherford.
Obtain Lise Meitner’s refined model and review it on the next page.**

Review of Lise Meitner's refined model:

	Not present	Present but incomplete			Present and complete
Present your refined model					
Describes macroscopic model	(1)	(2)	(3)	(4)	(5)
Describes molecular-level model	(1)	(2)	(3)	(4)	(5)
Molecular-level model is consistent with experimental evidence	(No)				(Yes)

Explain why your model has changed

Identifies key changes to macroscopic model	(1)	(2)	(3)	(4)	(5)
Identifies key changes to molecular-level model	(1)	(2)	(3)	(4)	(5)
Uses specific experimental evidence to explain changes to molecular-level model	(1)	(2)	(3)	(4)	(5)

Generalize your model

Provides generalized model that can be used to predict new situations	(1)	(2)	(3)	(4)	(5)
Uses specific experimental evidence to support generalized model	(1)	(2)	(3)	(4)	(5)

	Not at all accurate	Somewhat accurate			Completely accurate
How accurate do you believe your ratings above are?	(1)	(2)	(3)	(4)	(5)

Return Lise Meitner's model back to him/her.

Now obtain your own model once again and review it on the next page.

Review of Marie Curie’s own refined model:

	Not present	Present but incomplete			Present and complete
Present your refined model					
Describes macroscopic model	(1)	(2)	(3)	(4)	(5)
Describes molecular-level model	(1)	(2)	(3)	(4)	(5)
Molecular-level model is consistent with experimental evidence	(No)				(Yes)

Explain why your model has changed

Identifies key changes to macroscopic model	(1)	(2)	(3)	(4)	(5)
Identifies key changes to molecular-level model	(1)	(2)	(3)	(4)	(5)
Uses specific experimental evidence to explain changes to molecular-level model	(1)	(2)	(3)	(4)	(5)

Generalize your model

Provides generalized model that can be used to predict new situations	(1)	(2)	(3)	(4)	(5)
Uses specific experimental evidence to support generalized model	(1)	(2)	(3)	(4)	(5)

	Not at all accurate	Somewhat accurate			Completely accurate
How accurate do you believe your ratings above are?	(1)	(2)	(3)	(4)	(5)