2011

Argumentation in Undergraduate Chemistry Laboratories

Joi Phelps Walker
ARGUMENTATION IN UNDERGRADUATE CHEMISTRY LABORATORIES

By

JOI PHELPS WALKER

A Dissertation submitted to the
School of Teacher Education
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Spring Semester, 2011
The members of the committee approve the Dissertation of Joi Phelps Walker defended on March 3, 2011.

_______________________________________
Victor Sampson
Professor Directing Dissertation

_______________________________________
Kenneth Goldsby
University Representative

_______________________________________
Penny Gilmer
Committee Member

_______________________________________
Sherry Southerland
Committee Member

Approved:

_____________________________________
Lawrence C. Scharmann, Chair, School of Teacher Education

The Graduate School has verified and approved the above-named committee members.
I dedicate this dissertation to my mother, Dr. Pamela C. Phelps, who preceded me in this accomplishment and thus inspired me.
ACKNOWLEDGEMENTS

I came to graduate school with the question – how can I improve science teaching and measure the success or failure of the change? In keeping with the Argument Driven Inquiry instructional model, answering this question has been a collaborative effort. Over the last five years, a community of friends, colleagues and scholars has enabled me to pursue the answer to this question and to ultimately find my footing on the path to becoming a science education researcher.

In 1997, I was hired to teach chemistry at Tallahassee Community College (TCC) and began working with Carol Zimmerman, an accomplished chemistry educator. Four years ago, Carol and I set out to improve the general chemistry laboratory program. We originally expected that this would be a relatively minor project, but then we talked to Vic Sampson, learned about ADI and decided to reform the entire laboratory program. Carol’s willingness to take on this massive curricular reform demonstrated, on reflection, a degree of trust for which I am both humbled and grateful. Through each stage of the project, Carol’s willingness to push back and argue with me until we reached consensus has resulted in a general chemistry laboratory program that I believe will come to be a benchmark in undergraduate education.

My sincere thanks to my Dean, Dr. Frank Brown, who supported, facilitated and tolerated my pursuit of this degree, the research and the reform of the laboratory program in a manner that many down the road at the University found surprising for an administrator.

The laboratory manager, Alan Stuart-Tilley, was the unsung hero of the project, but his willingness to set-up multiple laboratories, to conduct investigations of his own into different aspects of the laboratory investigations under consideration and to tolerate my intense nature all are deeply appreciated.

I would like to acknowledge all the instructors who have been willing “to try” a new way of teaching and whose suggestions have improved the program. In particular, my thanks to Adonis McQueen for allowing me to use his laboratory sections for this research during which I essentially bossed him around for ten weeks. His good humor and forbearance are appreciated. Last and certainly not least, the students who worked so very hard are to be congratulated for their resilience and growth.

Over the years, I investigated numerous graduate programs and even applied to some, but nothing seemed quite right, until I talked with Penny Gilmer. She was of course a perfect
inspiration, being a chemist as well as an education researcher. Penny had such confidence in me and encouraged me to try science education for which I am in her debt. My second semester in the program I met Sherry Southerland and the die was cast. Sherry’s insight into teaching and learning opened doors that I had never known existed. The notion that we could study, evaluate and understand the processes of teaching and learning was the most exciting concept I had encountered since my sophomore organic chemistry course. To both of these exceptional women, I extend my utmost gratitude for their inspiration and support.

I am just glad that Ken Goldsby is teaching chemistry at the Florida State University. His honest joy in teaching chemistry, reminds me that ultimately we must love what we do if we are to have any hope of bringing students into the field. I thank Ken for being willing to step out of his comfort zone to think a bit about how we teach chemistry and to consider alternatives.

I have shared this journey with many students and I am indebted to each for their friendship and their guidance. Two dear friends, contributed significantly to this study and to my growth as a science educator, Patrick Enderle and Jonathon Grooms. Patrick and Jon scored data for me to ensure validity, helped me set-up assessments, ran video cameras, listened and advised. The importance of their role in this work was substantial and I thank them both sincerely.

To Vic Sampson, I can only relay a conversation I had recently with my daughter. Meg asked, “isn’t weird that Vic is younger than you?” I replied that one of the more remarkable aspects of my graduate school experience has been how much I could learn from someone 15 years younger than me. The ability to teach is for some a divine gift and that is the case with Vic Sampson. I was just lucky enough to be at the front of the line.

At the end of the line all too often in this journey was my beloved husband, Eric, and my dear daughters, Meg and Sarah. I have been asked how I could possibly work full-time, take classes and manage a family. The answer was simple; I had the full support of my family – thank you dear ones.

My sister, Rene’, has been a constant support and willing listener. Our husbands will be disappointed to find that we still have lots to talk about. Thanks dear friend.

I always knew that I would come home to Florida State; my parents were undergraduates when I was born and pulled me around campus in a wagon. My mother even took me to her child development class where I demonstrated my advanced skill with pop-beads. The circle is now complete.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS........................................................................................................... iv

LIST OF FIGURES ....................................................................................................................... ix

ABSTRACT.................................................................................................................................... x

CHAPTER 1 OVERVIEW OF STUDY................................................................................................. 1

Introduction.................................................................................................................................. 1

Theoretical Foundation................................................................................................................. 4

The ADI Instructional Model ....................................................................................................... 8

Previous Research on ADI.......................................................................................................... 14

Research Questions....................................................................................................................... 15

CHAPTER 2 METHODS AND METHODOLOGY ........................................................................ 16

Overview of the Research Design ................................................................................................. 16

Participants................................................................................................................................. 16

Instructional Procedure.............................................................................................................. 18

Data Sources ............................................................................................................................ 20

   Performance-based Assessment.......................................................................................... 20
   Assessment of Scientific Argumentation in the Classroom (ASAC).................................. 24
   Written Arguments............................................................................................................... 26

Data Analysis............................................................................................................................ 27

   Performance-based Assessment......................................................................................... 27
   Written Arguments............................................................................................................... 29

CHAPTER 3 RESULTS.................................................................................................................... 31

Introduction.................................................................................................................................. 31

   Quality of the Arguments Generated on the Performance-based Assessment ............. 31
   Quality of the Written Arguments Produced in Laboratory Reports......................... 33
   Quality of the Argumentation Episodes That Took Place During the Investigations.... 35
   The Relationship Between the Argumentation Events and the Written Arguments ..... 37

CHAPTER 4 DISCUSSION.................................................................................................................. 39
LIST OF TABLES

Table 1  Enrollment Demographics for General Chemistry Laboratory for Summer 2010........ 17
Table 2  Group Demographics for Each Section................................................................. 19
Table 3  Data Collection Timetable..................................................................................... 21
Table 4  Description of Performance-based Assessments ..................................................... 23
Table 5  Composite ASAC Scores for the Density Investigation for the Presenting Students .... 29
Table 6  Generation of Individual ASAC Scores for Density Investigation for the Traveling
         Students ........................................................................................................................................ 29
Table 7  Means and Standard Deviations for Performance-based Assessment Scores........... 31
Table 8  Means and Standard Deviations for Performance-based Assessment Component Scores
         ......................................................................................................................................................... 32
Table 9  Results of Paired Samples t-Tests for Performance-based Assessment Component Scores
         ......................................................................................................................................................... 33
Table 10  Means and Standard Deviations for Written Arguments........................................ 34
Table 11  Means and Standard Deviations for ASAC Scores (N = 9)...................................... 36
Table 12  Means and Standard Deviations Component ASAC Scores (N = 9)....................... 37
Table 13  Correlations among the Three Argument Scores (N = 95)...................................... 38
Table 14  Written Argument Scores for a Single Student ...................................................... 43
LIST OF FIGURES

Figure 1. Framework for a scientific argument used in ADI. ........................................... 7

Figure 2. The seven steps of ADI. ......................................................................................... 8

Figure 3. Three variations of the performance-based assessment. ................................... 23

Figure 4. Answer for one flask in the mid-term performance-based assessment. ............ 28

Figure 5. Example of written argument scoring. ............................................................... 30

Figure 6. Performance-based assessment scores. ............................................................... 32

Figure 7. Scores on each component of the performance task. ....................................... 34

Figure 8. Written argument scores. .................................................................................. 35

Figure 9. ASAC scores for both argumentation events. ................................................... 36

Figure 10. Scatter plot for argument evaluation scores vs. written argument scores. ....... 38
ABSTRACT

To address the need for reform in undergraduate science education a new instructional model called Argument-Driven Inquiry (ADI) was developed and then implemented in a undergraduate chemistry course at a community college in the southeastern United States (Sampson, Walker, & Grooms, 2009; Walker, Sampson, & Zimmerman, in press). The ADI instructional model is designed to give a more central place to argumentation and the role of argument in the social construction of scientific knowledge. This research investigated the growth in the quality of the student generated arguments and the scientific argumentation that took place over the course of a semester. Students enrolled in two sections of General Chemistry I laboratory at the community college participated in this study. The students worked in collaborative groups of three or four. The students were given a variation of the same performance task three times during the semester in order to measure individual ability to use evidence and justify their choice of evidence with appropriate rationale. Five ADI investigations took place during the semester and the laboratory reports for each were collected from each student and the argument section of each report was scored. All the student groups were video recorded five times during the semester as they generated and evaluated arguments and the quality of the group argumentation was assessed using an instrument called the Assessment of Scientific Argumentation in the Classroom (ASAC) observation protocol. As time was the independent variable in this study a repeated measure ANOVA was used to evaluate the significance of student improvement in each area (argumentation, written argument and performance task) over the course of the semester (Trochim, 1999). In addition, a multiple regression analysis was conducted to evaluate how well the ASAC scores predicted individual scores on both the performance task and the written arguments (Green & Salkind, 2005). There was significant growth over the course of the semester in all three measures, performance-based assessment, written argument and oral argumentation. There also was a significant correlation between written and oral arguments that was used to generate a linear model using oral argumentation as a predictor of written argument. The results of this suggest that the use of an integrated instructional model such as ADI can have a positive impact on the quality of the arguments students include in their investigation reports, the argumentation they engage in during lab activities, and their overall performance on tasks that require them to develop and support a valid conclusion with genuine evidence.
CHAPTER 1
OVERVIEW OF STUDY

Introduction

The current era of reform in science education can be traced to the publication of *A Nation at Risk* (National Commission on Excellence in Education, 1983). This U.S. government publication sounded the alarm on the economic and societal consequences of a failing education system. A cascade of reform documents followed ushering in the “standards movement” (American Association for the Advancement of Science, 1989; National Research Council, 1996). More significant perhaps to science education than the individual standards is the overarching goal of a scientifically literate society promoted by these documents. Scientific literacy is defined in the *National Science Education Standards* (NSES) as “the knowledge and understanding of scientific concepts and processes required for personal decision-making, participation in civic and cultural affairs, and economic productivity” (NRC, 1996, p. 22).

Achieving this goal has proven elusive for a many reasons, but certainly the disconnect between the goals of science literacy in the K12 system and undergraduate science education has played a significant role. Scientific literacy is not a focus of most undergraduate science courses, and it has not been prominent in the discourse of college science educators. Instead, undergraduate educators tend to focus on the goal of preparing students for disciplinary-based science careers. This objective requires an emphasis on disciplinary content and associated methodologies or processes. Undergraduate science students, as a result, tend to spend their early years learning the content of their disciplines and only later learn the associated processes through apprenticeship either in graduate school or in the workforce. Undergraduate science education rarely, if ever, gives undergraduate students an opportunity to learn about science as an epistemic endeavor (Abd-El-Khalick, 2001).

Two documents, *Shaping the Future: Strategies for Revitalizing Undergraduate Education* (National Science Foundation, 1996) and *Transforming Undergraduate Education in Science, Mathematics, Engineering, and Technology* (NRC, 1999), revisited many of the broad arguments and ideas developed in the K12 documents with a specific focus on undergraduate science and math education. These reports call for special attention to be given to introductory and lower-level science and math courses with the intent of restructuring the curricula to incorporate the connection of science and math with society and the human condition. The
authors of the reports support this vision for all undergraduate programs as a commitment to K12 teacher preparation and continuing the efforts of science education reform established at the K12 level. These national calls for reform have prompted some meaningful responses at various levels and institutions across the country.

A primary recommendation for science reform is for more inquiry-based instruction in science courses (NRC, 1996; 1999, 2000; NSF, 1996). Adapting the large lecture courses to an inquiry-based approach presents a considerable challenge, however the laboratory provides an ideal venue for this type instruction. Yet, traditional laboratory activities have been found to contribute little to student understanding of science or the development of scientific habits of mind (Burke, Greenbowe, & Hand, 2006; Cooper, 1994; Hofstein & Lunetta, 2004). Historically, undergraduate science laboratories served little purpose other than to allow students to demonstrate for themselves the concepts presented in a lecture and to perhaps provide some training in laboratory technique. Most laboratory curricula are therefore designed to ensure that students get the “right” results and draw the “appropriate” conclusions from these results. To accomplish this students are typically supplied with handouts that explain the premise behind the laboratory and the expected result. These handouts also include a procedure for collecting the data, a data table to facilitate data collection, and a step-by-step procedure for analyzing the data. Although these verification or “cookbook” labs may be highly efficient for introducing students to many different topics over the course of a semester, virtually no meaningful learning takes place when students participate in them (Domin, 1999; Hofstein & Lunetta, 2004). Students also tend to describe these types of labs as tedious, dull, and meaningless (Sunal, Wright, & Day, 2004). This type of lab experience coupled with a lecture course that often lacks relevance, relegates students to passivity, emphasizes competition, and focuses on algorithmic problem solving, has been faulted for the high student attrition in introductory science courses in general (Seymour, 2002; Tobias, 1992). In response to this finding, the Society of College Science Teachers (SCST) has recommended that laboratory courses be redesigned so the activities are more inquiry-based, grounded in current research on how people learn, and promote more critical thinking, problem solving, and collaborative work on meaningful tasks (Halyard, 1993). Whether the goal is to retain students in science programs or to advance the goals of the NSES and Transforming Undergraduate Education, the suggested solution is to move away from tightly directed experiments meant to confirm some obvious principle toward a more authentic
science experience (NRC, 2000; Siebert & McIntosh, 2001; Sunal et al., 2004; Tobias, 1992).

To address the need for reform in undergraduate science education a new instructional model called *Argument-Driven Inquiry* (Walker, Sampson, & Zimmerman, in press) has been developed and implemented into a undergraduate chemistry course at a community college in the southeastern United States. This instructional model is grounded in social constructivist theories of learning (Driver, Asoko, Leach, Mortimer, & Scott, 1994; Vygotsky, 1978) and is designed to make laboratory experiences more scientifically authentic and educative for students (Anderson, 2007; Bransford, Brown, & Cocking, 1999; Driver et al., 1994). The laboratory activities are more authentic in nature because they closely mimic those of science research laboratories. These activities are also more educative for students because they receive feedback throughout the process and have an opportunity to learn from their mistakes. Laboratory experiences designed using the Argument-Driven Inquiry instructional model consist of seven phases or steps that require students within a collaborative group to develop a procedure or process for answering a given research question, followed by production of an initial argument that articulates and justifies an explanation for the phenomenon under investigation (argument). Multiple groups then engage in an argumentation session in order to share their arguments with other students and critique the work of others in order to reach consensus. These first steps are completed in a single laboratory session. The students then write an investigation report on their own that are submitted the following week to a double-blind peer-review that is also a collaborative group activity. Based on results of the peer review students submit a revised report to the instructor during the third week of the sequence. These steps have been intentionally designed to provide students with a richer and more genuine science laboratory experience and one that provides students with exposure to a community of practice that closely resembles the scientific community.

The overall goal of this study was to examine how a group of undergraduate students’ ability to participate in collaborative scientific argumentation and to craft a written argument changes over the course of a semester as they completed a series of laboratories designed using the ADI instructional model. Most participants initially demonstrated minimal competence in both collaborative argumentation and individual construction of a scientific argument. As the semester progressed and students gained experience the quality of their arguments improved as did their ability to engage in the process of argumentation. The extent to which there was transfer
from oral discourse to written argument was evaluated as well.

**Theoretical Foundation**

The dominant method of instruction in undergraduate science has, and continues to be, a teacher-centered lecture based model that is designed to be efficient and cost-effective for large-scale information delivery (Spencer, 1999; Sunal et al., 2004). However, the lecture hall is usually not the place where students develop an understanding of science content, rather informal study groups, recitation or individual instruction with the professor often fulfill this purpose (Wamser, 2006). The traditional lecture format used in undergraduate education is based on a transmission theory of learning. This perspective is founded on the assumption that ideas can be transferred intact from the mind of the instructor to the mind of the student (Bransford et al., 1999; Spencer, 1999). The failure of this model to produce not only students with the expected content knowledge but the ability to solve problems or think critically within a given discipline is the focus of national documents as well as others calling for reform in undergraduate education (NRC, 1999; NSF, 1996; Sunal et al., 2004; Taylor, Gilmer, & Tobin, 2002). In addition, the emphasis on ‘final form’ science without connections to the process of science undermines efforts for science literacy as described in the introduction.

Current understandings of how people learn are based on constructivist theories of learning (Bransford et al., 1999; Tobin, 1993; von Glasersfeld, 1989). Constructivist theories of learning are founded on the basic assumption that knowledge cannot be transmitted directly from one person to another, but is actively built or “constructed” by the learner (Driver et al., 1994). This construction of knowledge can occur individually (personal constructivism) or as a social process (social constructivism). Personal constructivism describes the process by which individuals develop their own meaning or sense of the natural world through individual processes, a theory first described by Piaget (1964; 2003). Piaget described learning as a process of assimilation and accommodation of new information based on what an individual already knows. Thoughts concerning the social construction of knowledge emerged from the work of Vygotsky (1978), who described how the interactions that occur amongst groups of learners influence the nature of the knowledge that is constructed by an individual. In science these interactions are embedded within a culture that is shaped by tools and symbols which are not found in nature, but are constructs of the scientific community (Hodson, 2008). It is therefore important to focus on how individuals learn to use the tools, symbols, and language of the
scientific community (Driver et al., 1994) and how the tools, symbols, and language of the scientific community shape what an individual learns.

In order to teach science from a social constructivist perspective, however, requires a reconsideration of how a laboratory course can be designed to help foster learning (Driver, Newton, & Osborne, 2000). In light of social constructivist views of learning, rather than being a place where students verify concepts taught in the lecture, the laboratory should provide a setting where students can interact directly with the material world, using the tools, data collection techniques, models and theories of science (NRC, 2006). The laboratory, it can be argued, is the ideal setting for accessing the aspects of science that may be missing in the lecture. For example, students should conduct investigations in conversation and collaboration with peers as well as more experienced adults. Through these types of interactions, learners encounter and explore ideas in a subject, use for these ideas, and ways that ideas are validated; that is, students learn what constitutes legitimate knowledge in a field (Linn & Burbules, 1993). Ford (2008) calls this a “grasp of practice” and argues that it is fundamental to learning science. The specific disciplines that underlie the undergraduate level courses, such as Biology and Chemistry, have established bodies of knowledge, a unique language, and rules for gathering evidence and evaluating results. As students engage in conversations with others, they can draw on their expertise; explain, extend, and reflect on their own ideas; and gain exposure to the ways of thought exhibited by disciplinary experts (Sunal et al., 2004).

To bring about such reforms in science teaching requires changes to be made to the roles student must assume, the nature of content and tasks, assessments, and the social organization of the laboratory. Such changes entail numerous challenges. One involves creating new pedagogies that reflect these perspectives on learning. It should therefore not be surprising that over the last 20 years a number of developers and researchers have created new approaches to science education that are more consistent with the tenets of social constructivism (e.g., Farrell, Moog, & Spencer, 1999; Grosser et al., 1996; Lewis & Lewis, 2008; Poock, Burke, Greenbowe, & Hand, 2007). One such approach is called Argument-Driven Inquiry or ADI (Walker, Sampson, & Zimmerman, in press). This instructional model is designed to give a more a central place to argumentation and the role of argument in the social construction of scientific knowledge and the interpretation of empirical data during laboratory-based instruction. In the sections that follow, I will discuss the ADI instructional model in more detail. However, before I can discuss the
nature of ADI, I will first need to clarify the meaning of the terms argument and argumentation and to elaborate on the current research in this area.

In science, argumentation is not the societal idea involving conflict and disagreement; rather it is a form of “logical discourse whose goal is to tease out the relationship between ideas and evidence” (Duschl, Schweingruber, & Shouse, 2007, p. 33). Scientific argumentation, as a result, plays a central role in the development, evaluation, and validation of scientific knowledge and is a core epistemic practice (way of knowing) that distinguishes science from other disciplines (Bricker & Bell, 2008; Driver et al., 2000; Duschl & Osborne, 2002). Instructional approaches that provide opportunities for students to engage in argumentation, therefore, not only provide a venue for the social construction of scientific ideas but can also provide a chance for students to improve their understanding of the scientific enterprise as well (Duschl & Osborne, 2002; Linn & Burbules, 1993).

In this type of research, it is important to define and delineate between the product and the process of argumentation. An argument refers to the actual claim and associated reasons provided in support of the conclusion and argumentation focuses on the social processes involved in generating those arguments (Erduran, Simon, & Osborne, 2004; Kuhn & Udell, 2003; Sampson & Clark, 2006). A scientific argument (as represented in Figure 1) consists of a claim supported by evidence and other reasons. The rationale for the scientific argument explains not only why the evidence supports the claim but also establishes the validity of the evidence as well.

Scientific argumentation requires students to analyze and evaluate data and then rationalize its use as evidence for a claim. The evidence component of an argument refers to measurements or observations gathered by the students that are used to support the validity or the acceptability of the conclusion. This evidence can take a number of forms ranging from traditional numerical data (e.g., pH, mass, temperature) to observations (e.g., color, descriptions of an event). However, in order for this information to be considered evidence it needs to either be used to show (a) a trend over time, (b) a difference between groups or objects, or (c) a relationship. The unifying component of an argument is the rationalization that explicates why the evidence supports the conclusion and why the evidence provided should count as evidence.
Figure 1. Framework for a scientific argument used in ADI.

The concern with argumentation is not that students cannot support ideas and challenge claims or viewpoint when debating everyday issues (e.g., Baker, 1999; Pontecorvo & Girardet, 1993; Resnick, Salmon, Zeitz, Wathen, & Holowchak, 1993; Stein & Miller, 1993). Instead, it is within the context of science that students’ argumentation skills fall short (Driver et al., 2000). A number of studies, for example, have shown that students will manipulate, distort or trivialize data rather than reject a preconceived conclusion (Berland & Reiser, 2009; McNeill & Krajcik, 2007; Sandoval & Millwood, 2005). Other researchers, who have examined the ways students craft a written argument, have found that students are unable to discern what data are relevant or what counts as evidence (Kelly & Bazerman, 2003; Kelly & Takao, 2002). Instead, students tend to provide details about their methods or observations rather than using appropriate evidence and reasoning to support their claim (Zeidler, 1997). Finally, even when students have grasped the concept of evidence they will submit the evidence as all inclusive, neglecting to rationalize the use of this evidence in supporting their conclusion or claim (e.g., Klahr, Dunbar, & Fay, 1990; Schauble, Klopfer, & Raghavan, 1991). The available literature, in summary,
suggests that students’ lack of argumentation skills in science reflects their misunderstanding of ‘what counts’ in science rather than being indicative of low cognitive abilities or poor general argumentation skills.

**The ADI Instructional Model**

The ADI instructional model is designed to give a more a central place to argumentation and the role of argument in the social construction of scientific knowledge while promoting inquiry. Inquiry and argumentation are complementary goals that make laboratory experiences more scientifically authentic and educative for students (Jimenez-Aleixandre, 2007). The ADI instructional model consists of seven components or steps (see Figure 2). Each step of the model is equally as important as the next in successfully achieving the intended goals and outcomes of the process. All seven stages are therefore designed to be interrelated and to work in concert with the others. In the paragraphs that follow, I describe each of these steps and provide justification.
for inclusion of each in the instructional model.

The first step of the ADI instructional model is the identification of the task through discussion of the research question. Similar to other instructional models, such as the Science Writing Heuristic (Wallace, Hand, & Yang, 2005) or the Cooperative Chemistry Lab Model (Cooper, 1994), this step is designed to provide students with a challenging problem to solve. It is also designed to capture the students’ attention and interest. This first step is important because studies have shown that “students often perceive that the principal purpose for a laboratory investigation is either following the instructions or getting the right answer” (Hofstein & Lunetta, 2004, p. 38). This perception is likely due to the nature of the tasks in traditional laboratory methods that entail either a process to be practiced or a value to be measured. For example, *Separation of Salt and Sand* or *Determination the Ideal Gas Constant* are both lab activities found in a typical laboratory manual (Dingledy et al., 2004), which focus on following procedures rather than answering questions. To present the laboratory investigation as an opportunity to discover something or solve some problem, a good research question is critical. The research question provides the foundation for scaffolding student argumentation and creating a need for evidence. For example, a beginning ADI investigation provides the group of students with a set of objects and asks, “Are these objects made of the same material?”

The second step of the ADI instructional model focuses on the generation of data. In this step of the model, students work in a collaborative group (3 or 4 students) in order to develop and implement a method (e.g., an experiment or analysis) to answer the research question provided in step 1. A collaborative group is defined as a group of students working in concert to accomplish a common goal (Linn & Burbules, 1993). The intent of this step is to provide students with an opportunity to learn how to design and conduct informative investigations and to learn to deal with the ambiguities of empirical work. The nature of these investigations is best described as a “guided inquiry” because each group of students must decide the ways to gather and analyze the data they will need to justify an answer to the research question (Bell, Smetana, & Binns, 2005). This step is important because many undergraduate students never have an opportunity to learn how to develop their own methods to investigate a topic or to answer a research question. Instead, students are typically given detailed directions to follow and then “gather and record data without a clear sense of the purposes and procedures of their investigation and their interconnections” (Hofstein & Lunetta, 2004, p. 40).
The third step in the ADI instruction model is the *production of a tentative argument*. This component of the instructional model calls for students to craft a written argument that consists of an explanation supported with evidence, and a rationale for the choice of evidence in a medium, such as a large whiteboard, that can be shared with others. This step is included in the model because developing assertions about the natural world inside the laboratory and then justifying those assertions with data collected during an investigation are considered by many to be an important element of science learning (Driver et al., 2000; Duschl, 2008; Lunetta, Hofstein, & Clough, 2007). The literature that outlines the current goals for science education at the undergraduate level (AAAS, 1989; NRC, 1996; 1999, 2000; NSF, 1996), also support this notion. These documents call for more opportunities for learners to use data and current scientific knowledge to construct explanations “based on evidence and logical argument” (NRC, 1996, p. 145). In addition, current research indicates that these written artifacts provide a venue to promote and support the critical evaluation of ideas inside the science classroom and can stimulate more personal reflection about the topic (Duschl, 2007; Kelly, Regev, & Prothero, 2007; Sandoval & Reiser, 2004).

This third step of the model is designed to emphasize the importance of argument in science. In other words, students need to understand that scientists must be able to support an explanation, conclusion, or other claim with appropriate evidence and a rationale because scientific knowledge is not dogmatic (Hodson, 2008). This third step is also included to help students develop a basic understanding of the elements that count as a high quality argument in science and ways to determine if available evidence is valid, relevant, sufficient, and convincing enough to support a conclusion (or claim). More importantly, this step is designed to make students’ ideas, evidence, and rationale visible to each other; which, in turn, enables students to evaluate alternatives and eliminate conjectures or conclusions that are inaccurate or do not fit with the available data (something done in the next stage of the instructional model).

The fourth step in the ADI instructional model is an *argumentation session*. During an argumentation session, the small groups have an opportunity to share their arguments and to critique the arguments of other groups. It also gives the students a chance to evaluate alternative conclusions in order to determine which is the most valid or acceptable or to refine a conclusion to make it more valid or acceptable. In other words, argumentation sessions are designed to give students an opportunity to learn to critique the products (i.e., conclusions, explanations or
arguments), processes (i.e., methods), and context (i.e., theoretical or empirical foundations) of an inquiry. This step is included in the model because research indicates that students learn more when they are exposed to the ideas of others, respond to the questions and challenges of their classmates, support their ideas with evidence and other reasons, and evaluate the merits of alternatives (Linn & Eylon, 2006; NRC, 2007). It also provides teachers with an opportunity to assess student progress or thinking and to encourage students to think about issues that may have been overlooked or ignored.

The argumentation sessions are intended to promote and support learning by taking advantage of the variation in student ideas that are found within a classroom. It also gives students a chance to negotiate meaning (Hand, Norton-Meier, Staker, & Bintz, 2009) and to adopt new standards for evaluating claims in the context of science. The ability to evaluate claims in the context of science is important because research suggests that students often rely on criteria such as plausibility or external authority in order to determine which ideas to accept or reject as they attempt to negotiate meaning with others (Kuhn & Reiser, 2005; Sampson & Clark, 2009). The argumentation sessions are included in the model in order to help students learn to use criteria valued in science, such as consistency with scientific theories or laws or fit of data, to distinguish between alternative conclusions or explanations. It also gives students an opportunity to refine and improve on their initial ideas or methods.

The fifth step of the ADI instructional model is the creation of a written investigation report by individual students. This report gives the students an opportunity to share the goal of their investigation, the method they used, and their overall argument. This step is included in the model for several reasons. Requiring students to write helps them make sense of the topic because the writing process encourages metacognition (Wallace, Hand, & Prain, 2004). As a result, an opportunity to write can actually help students learn and retain content (Indrisano & Paratore, 2005). In addition, writing an investigation report as part of a laboratory activity can provide a venue for students to learn “the disciplinary knowledge, norms, and practices” that make science different from other ways of knowing (Kelly et al., 2007, p. 139).

Writing an investigation report, however, can be difficult for students to complete without guidance. In order to craft a high quality written argument, individuals need to understand how to make evidence clear to the audience and how to make a claim appear to be valid or acceptable (Bazerman, 1988; Kelly & Bazerman, 2003; Kelly & Takao, 2002; Latour, 1987). Unfortunately,
most students lack the content knowledge and the skills needed to write well in science (Kelly et al., 2007). The ADI instructional model therefore uses a non-traditional report format in order to help students learn to write in science and to help them better understand the content. The intent of this format is to bring the persuasive aspects of science writing to the foreground and to highlight the non-narrative structure (i.e., text that is hierarchical or clustered by topic) and the multi-modal presentation of information (e.g., words, figures, tables, equations, etc.) that characterize science texts (Bazerman, 1988; Carter, 2007; Wallace et al., 2004). This report format is also designed to encourage students to think about ‘what they know’, ‘how they know it’, and ‘why others should accept a given explanation over alternatives.’

To help students learn to write in science, ADI requires students to produce a report that is organized into three sections around three fundamental questions: What were you trying to do, and why? What did you do, and why? What is your argument? Students are also encouraged to organize the data they gathered and analyzed during the second step into tables or graphs that they embed into the report and reference in the body of the text in order to help students learn to communicate information in multiple modes. The three questions students address as they write target the same information that is included in more traditional laboratory report formats (e.g., introduction, procedure, results, and discussion) but are designed to help students understand the importance of argument in science, to elicit their awareness of the non-narrative text structure and audience, and to help the make sense of the content (e.g., ‘what they know’ and ‘how they know’) as they write.

The sixth step of the ADI instructional model is a double-blind peer-review of these reports to ensure quality. Once students complete their investigation reports they submit four typed copies identified only by a code number assigned by the classroom teacher. The teacher then randomly distributes three or four sets of reports (i.e., the reports written by three or four different students) to each laboratory group along with a peer review handout for each set of reports. The peer review handout includes specific criteria to be used to evaluate the quality of an investigation report and space to provide feedback to the author. The review criteria include questions such as: Did the author make the research question and/or goals of the investigation explicit? Did the author describe how they went about his or her work? Did the author use genuine evidence to support their explanation? Is the author’s reasoning sufficient and appropriate? The lab groups review each report as a team and then decide if it can be accepted as
is or if it needs to be revised based on the criteria included on the peer review sheet. Groups are also required to provide explicit feedback to the author about what needs to be done in order to improve the quality of the report and the writing as part of the review.

This sixth step of the instructional model is designed to provide students with educative feedback, encourage students to develop and use appropriate standards for ‘what counts’ as quality, and to help students be more metacognitive as they work. It is also designed to create a community of learners that values evidence and critical thinking inside the classroom. This is accomplished by creating a learning environment in which students are expected to hold each other accountable for the quality of their conclusions and arguments. Students, as a result, are taught to expect to discuss the validity or the acceptability of conclusions as part of each investigation and begin to adopt more and more rigorous criteria for evaluating or critiquing them. This type of focus also gives students a chance to see both “strong” and “weak” examples of investigation reports written for the same purpose. Students must be able to read, understand, and evaluate the quality of science writing to be literate in science. In order for students to do this, they need to learn about ‘what counts’ as quality in the context of science (Shanahan, 2004).

The seventh, and final, step of the ADI instructional model is the revision of the investigation report based on the results of the peer-review. The reports that are accepted by the reviewers may be submitted to the instructor at the end of step 6; however, all students (even when their first draft is “accepted as is”) have the option to revise their reports based on what they have read and the comments on their initial draft. Authors who wrote papers that were not accepted by their peers, on the other hand, are required to rewrite their reports based on the reviewers’ comments and suggestions. Once completed, the revised reports (along with the original version of the report and the peer review sheet) are submitted to the instructor for a final evaluation. This approach is intended to provide an opportunity for students to improve their mechanics, reasoning, and their understanding of the content without imposing a grade-related penalty. It also provides students with a chance to engage in the entire writing process (i.e., the construction, evaluation, revision, and eventual submission of a manuscript) in the context of science. This step, once completed, brings the instructional sequence to a close and provides the instructor with an authentic assessment of the students’ understanding of the content and their ability to formulate a scientific argument.

The ADI instructional model, in summary, is designed to function as a short integrated
instructional unit (NRC, 2006) and to encourage students to engage in a sequence of activities (inquiry, argumentation, writing, and peer review) that are intended to help students understand important concepts and practices in science. For example, engaging in argumentation requires individuals to make sense of data, generate and articulate explanations for natural phenomena, justify explanations with appropriate evidence and reasoning, and critique the validity of alternative viewpoints (Driver et al., 2000; Duschl, 2007; Kuhn, 1993; Osborne, Erduran, & Simon, 2004; Sampson & Clark, 2009). Writing requires students to be able to articulate their thinking in a clear and concise manner, encourages metacognition, and makes students’ thinking and reasoning visible to others (Wallace et al., 2004). Peer-review provides students with educative feedback (Wiggins, 1998), encourages students to develop and use appropriate standards for ‘what counts’ as quality (Sandoval & Reiser, 2004), and can help students be more metacognitive as they work (NRC, Donovan & Bransford, 2005; 2007). A number of studies indicate that such activities can help students learn content (Bell & Linn, 2000; Driver et al., 1994; Zohar & Nemet, 2002), develop complex reasoning and critical thinking skills (Driver et al., 2000; Kuhn, 1992, 1993; Lawson, 2003), and improve communication skills (Kuhn & Udell, 2003; Wallace et al., 2005).

**Previous Research on ADI**

The ADI instructional model was piloted in select sections of General Chemistry I laboratories during the 2008-2009 academic year. During these three semesters the remaining laboratory sections were taught using a traditional approach in order to compare the achievement of the students who were enrolled in the ADI sections to the students in the traditional lab sections (Walker, Grooms, Anderson, Zimmerman, & Sampson, 2010; Walker, Sampson, Grooms, Zimmerman, & Anderson, in press). The results of this study indicated that the students in the ADI sections had an equivalent conceptual understanding of key concepts (as measured with conceptual inventory) as students in the traditional sections despite completing fewer laboratory investigations. Students in the ADI sections were better able to provide evidence and reasoning to support their conclusions (in both familiar and unfamiliar contexts). The female students in the ADI sections also had significantly better attitudes towards science at the end of the semester when compared to the female students in the traditional sections of the course.

In a separate, but related study, the extent that the ADI instructional model influenced science writing was also explored (Sampson, Walker, Dial., Swanson, & Zimmerman, in
review). In this study, Sampson et al. (in review) made copies of the initial and final version of each report and the peer review guide completed by each group for each investigation and the instructor scoring sheet and then scored the reports written by the students with the same rubric. The results of this study, although exploratory, suggest that both strong and weak writers can make substantial gains in their science writing skills over the course of a semester with ADI.

**Research Questions**

The results of these two studies, when taken together, provided the justification needed to implement the ADI instructional model into all the laboratory sections and for continued research on the influence of the ADI instructional model on student learning. There are numerous directions that this research could take, ranging from investigation of students’ understanding of the nature of science, improvement of inquiry skills, or the nature of student collaborative work. The aspect of research on the ADI instructional model that I focused on was the process of argumentation in collaborative groups and the crafting of an argument by individuals. The following research questions guided this study.

a) To what extent does the ADI instructional model improve undergraduate students’ performance on a task that requires them to develop and support a valid conclusion with appropriate evidence and an adequate rationale?

b) To what extent does the ADI instructional model improve the quality of the scientific argumentation that takes place between undergraduate students over the course of a semester?

c) To what extent does the ADI instructional model improve the quality of the scientific arguments undergraduate students write over the course of a semester?

d) To what extent does the quality of the scientific argumentation that takes place between undergraduate students predict individual undergraduate students’ written arguments?
CHAPTER 2
METHODS AND METHODOLOGY

Overview of the Research Design

This study was a naturalistic inquiry (Denzin & Lincoln, 2003) into how a group of learners responded to the ADI instructional model. I focused specifically on how participants’ arguments and argumentation changed over time and on the conditions that influenced these changes. To help me characterize the influence of various instructional experiences on the participants’ thinking, I attempted to track participants’ performance using three types of assessment: a performance-based assessment given at three points in the semester, arguments written by each student for each laboratory activity, and the argumentation that took place within the groups during each of the laboratory activity. These data sources were used to measure the impact of the ADI instructional model on growth in student argumentation in groups and individually over the course of the semester. The intent of the longitudinal study was to evaluate more than just absolute change from beginning to end. Rather by looking at multiple measures over the course of the semester I was able to capture the process of change that occurred for the collaborative groups as well as the individual students. In addition, significant correlations between the written and oral argument enabled me to ascertain the degree to which group performance impacted individual written arguments.

Participants

Students enrolled in two sections of General Chemistry I Laboratory at a community college in the southeastern United States participated in this study. The community college is an open admission, comprehensive two-year institution that serves nearly 40,000 students annually at a main campus and four service centers located throughout the college’s statutory service area. All the laboratory sections for General Chemistry I were taught using the ADI instructional model. This study was conducted during the summer semester which at the community college lasts ten weeks rather than the 15 weeks of a typical semester. This resulted in an accelerated pacing for the laboratory course such that students were in laboratory for four hours some days, essentially completing two weeks’ worth of material relative to the longer semester pacing.

An adjunct instructor with two semesters of experience using the ADI instructional model in the chemistry laboratory taught the laboratory sections used in this study. The two laboratory sections met at 2:30 pm on different days. There were originally 23 students enrolled in each
section. Over the course of the semester seven students withdrew, five from section A and 2 from section B. Table 1 presents demographic data for the Summer 2010 semester. The research participants were representative of the population in terms of gender, race, age and GPA.

Table 1

*Enrollment Demographics for General Chemistry Laboratory for Summer 2010*

<table>
<thead>
<tr>
<th>Enrollment</th>
<th>Female</th>
<th>Minority</th>
<th>Age</th>
<th>GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Range</td>
<td>Mean</td>
</tr>
<tr>
<td>All Sections</td>
<td>119</td>
<td>49%</td>
<td>47%</td>
<td>17 – 51</td>
</tr>
<tr>
<td>Participants</td>
<td>46</td>
<td>56%</td>
<td>44%</td>
<td>19 – 35</td>
</tr>
</tbody>
</table>

*Note: Minority means Black, Hispanic, Asian or Multiracial.*

The laboratory had six octagonal tables that accommodated up to four students. During the first laboratory meeting, the students were allowed to self-associate, but at the beginning of the second laboratory meeting students were assigned to a core group for the duration of the semester. The students were divided into groups by sorting a spreadsheet first by gender, then by GPA and assigning a table number (1 - 6) twice for female students and twice for male students. This resulted in six groups made up of both male and female students, a range of GPAs and of mixed race. Section B had four students that originally did not sign the consent form and they were placed together at Table 6. On the first day of filming, these students asked questions that alleviated anxiety they had regarding the research; they signed the consent form and were included in the study.

The community college setting presents a research context that is rich in diversity as shown in Table 1 and in the composition of the groups listed in Table 2. This student population, however, is not residential, many students work full or part-time jobs, they have families and other obligations that often hinder their education goals. These factors made data collection a challenge, due to attrition, absences, missed assignments or assignments that represent a minimal effort. As an example, during this study students in one of the laboratory sections were falsely led to believe that laboratory had been cancelled one week and no one came to class. After some debate we were able to continue the research in this section, but just one week off from the other
section. Throughout the semester, students had personal issues that impacted their attendance or performance on assignments. Despite these contextual limitations, the ADI instructional model was implemented with overall fidelity in the two sections and the data collected can be considered representative of student work in an undergraduate chemistry laboratory.

I was present at all but one laboratory meeting as a participant observer. I kept a field journal, operated the video cameras, and took photographs. There was often a fellow graduate student present as well to assist with the video cameras. Students did turn to me for guidance during the labs and I had three approaches 1) signal the instructor to step in 2) suggest they check with other members of their group and 3) provide minimal assistance.

**Instructional Procedure**

The General Chemistry I Laboratory was intended to provide students with the opportunity to conduct investigations into concepts identified by the faculty as important for understanding chemistry. These concepts include physical properties, molecular structures, solutions, chemical reactions, and energy of reactions. These topics are listed in the course description for the State University System and are required for common course numbering (Education, 2009). The topics are not unique to the community college curriculum; rather it is the instructional method that distinguishes these laboratory activities. The previous mode of laboratory instruction was based on commercially available lab handouts designed to “deliver” a lab activity to a large number of students. The handouts told the students the premise behind the laboratory, the procedure for collecting the data, step-by-step procedure for the calculations, and finally the expected result. The ADI laboratory handouts (see Appendix A) provide a brief introduction to the topic, a statement of the problem to be investigated, the guiding question for the investigation, the materials available for use, safety precautions and getting started suggestions. The students are encouraged throughout the semester to keep a notebook in order to record the process they use to answer the guiding questions and to create data tables that are meaningful for their investigation.

The sequence of ADI investigations progressed from simple manipulation of glassware and objects (Density) to more advanced chemistry investigations (Chemical Reactions). The specific order of the ADI investigations is synchronized with the order of concept presentation in the lecture course. The first laboratory activity, ‘Using Laboratory Equipment and Glassware’, was not an ADI lab, but was used as an orientation to the laboratory and to introduce the students to
### Table 2

**Group Demographics for Each Section**

<table>
<thead>
<tr>
<th>Table</th>
<th>Section A</th>
<th></th>
<th>Section B</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age</td>
<td>Race</td>
<td>Gender</td>
<td>GPA</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>White</td>
<td>Female</td>
<td>2.50</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Black</td>
<td>Male</td>
<td>1.56</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Black</td>
<td>Female</td>
<td>2.92</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>White</td>
<td>Male</td>
<td>3.27</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Black</td>
<td>Male</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Hispanic</td>
<td>Female</td>
<td>2.09</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>White</td>
<td>Male</td>
<td>2.70</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Black</td>
<td>Female</td>
<td>2.28</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Black</td>
<td>Female</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Hispanic</td>
<td>Female</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Black</td>
<td>Male</td>
<td>3.00</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>White</td>
<td>Female</td>
<td>3.30</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>White</td>
<td>Male</td>
<td>2.37</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Hispanic</td>
<td>Male</td>
<td>2.94</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>White</td>
<td>Female</td>
<td>3.05</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Black</td>
<td>Female</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Black</td>
<td>Female</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Hispanic</td>
<td>Female</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Black</td>
<td>Female</td>
<td>2.43</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>White</td>
<td>Female</td>
<td>2.72</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Hispanic</td>
<td>Male</td>
<td>2.81</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Black</td>
<td>Male</td>
<td>3.45</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>White</td>
<td>Female</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Black</td>
<td>Male</td>
<td>2.22</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>White</td>
<td>Female</td>
<td>2.65</td>
<td></td>
</tr>
</tbody>
</table>

NA – current GPA was not available.

...some of the components of the ADI instructional model. The next five activities, Density, Hydrate, Solutions, Mole Ratio and Chemical Reactions, were ADI investigations that utilized all seven steps of the instructional model. The Hot Pack – Cold Pack activity was an abbreviated ADI investigation. The entire class conducted the investigation with each group investigating different sets of salts. There was no laboratory report for this investigation; instead students submitted calculations and an argument for the selected commercial product. Only the first four
steps of the ADI instructional model were utilized in this investigation. The final laboratory utilized a molecular model kit for investigating shapes of molecules and did not follow the ADI system.

**Data Sources**

During the first meeting of the laboratory sections the students were given the performance task to complete individually. This assessment established a baseline for individual argument skill as well as knowledge of limiting reactants. For each of the five ADI investigations the groups were observed until they began constructing their argument and designing their poster at which point video recording using a separate video camera for each table began and continued through the argumentation session and follow-up discussion. The posters for each group were photographed at the end of the argumentation session as well. The laboratory reports were copied and evaluated using an established rubric. Variations of the performance task were administered near the middle and at the end of the semester to capture individual growth over the semester. Table 3 outlines the timetable for data collection. I was present for all but one laboratory meeting as a participant observer and kept a journal of my observations. This enabled me to ensure fidelity of the instruction and monitor any individual issues that developed during the semester.

**Performance-based Assessment**

As science teaching methods expand to include more opportunities for students to develop critical thinking skills and investigative competency, or as in the case of ADI, learn ways to participate in argumentation, then assessments must be employed that measure multiple aspects of student achievement (Siebert & McIntosh, 2001). In this research I wanted to capture student growth in the ability to use evidence to justify a claim as supported by the ADI instructional methodology. To achieve this goal I developed a performance task to evaluate student understanding of limiting reactants in a simple chemical reaction (Walker, Sampson, Grooms, & Zimmerman, in press). Performance-based assessments such as this allow students to demonstrate their inquiry and reasoning skills as well as their understanding of the content (Doran, Chan, & Tamir, 1998). In addition, performance-based assessments connect and interface with the NSES (NRC, 1996), which argue that the content and form of an assessment task must be congruent with that which is being measured. The NSES further argue that assessment tasks that are similar in form to tasks in which students will engage in their lives.
outside the classroom or are similar to the activities of scientists are “authentic” and therefore more legitimate (NRC, 1996). While the NSES are written for K12 education, their application in undergraduate education is seen in expanding methods of instruction coupled with alternative assessments, such as concept mapping, journals, portfolios and performance-based assessments that require students to evaluate situations (Siebert & McIntosh, 2001).

Table 3

Data Collection Timetable

<table>
<thead>
<tr>
<th>Week</th>
<th>Activities</th>
<th>Data Collected</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction to ADI</td>
<td>Demographics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Performance Task #1</td>
</tr>
<tr>
<td></td>
<td>How to use Laboratory Equipment</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Density Investigation (Steps 1-4)</td>
<td>Video Recording (Steps 3 &amp; 4)</td>
</tr>
<tr>
<td>3</td>
<td>Density Investigation (Step 6)</td>
<td>Video Recording</td>
</tr>
<tr>
<td></td>
<td>Hydrate Investigation (Steps 1-4)</td>
<td>Investigation Report #1</td>
</tr>
<tr>
<td>4</td>
<td>Hydrate Investigation (Step 6)</td>
<td>Observations</td>
</tr>
<tr>
<td></td>
<td>Dilution Exercise</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Dye Solutions Investigation (Steps 1-4)</td>
<td>Video Recording (Steps 3 &amp; 4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Investigation Report #2</td>
</tr>
<tr>
<td>6</td>
<td>Dye Solutions Investigation (Step 6)</td>
<td>Performance Task #2</td>
</tr>
<tr>
<td></td>
<td>Mole Ratio Investigation (Steps 1-4)</td>
<td>Video Recording (Steps 3 &amp; 4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Investigation Report #3</td>
</tr>
<tr>
<td>7</td>
<td>Mole Ratio Investigation (Step 6)</td>
<td>Video Recording</td>
</tr>
<tr>
<td></td>
<td>Chemical Reactions Investigation (Steps 1-4)</td>
<td>Investigation Report #4</td>
</tr>
<tr>
<td>8</td>
<td>Hot Pack/Cold Pack Investigation</td>
<td>Observations</td>
</tr>
<tr>
<td>9</td>
<td>Molecular Shapes</td>
<td>Investigation Report #5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Performance Task #3</td>
</tr>
</tbody>
</table>
Common chemical demonstrations are easily adapted for performance-based assessments that can capture not only a student’s conceptual understanding but their ability to rationalize their use and choice of evidence as well (Doran et al., 1998). The performance task used in this research was based on the “Balloon Race” demonstration (Cesa, 2002). The “Balloon Race” demonstration entails mixing constant volumes of 1M acetic acid with five different amounts of sodium bicarbonate added from a balloon covering the mouth of a flask. The balloon inflates to specific volume based on the amount of sodium bicarbonate it contained. Adding universal indicator to the flasks provides an additional variable to consider as the solutions that are initially acidic either retain their acidity or become basic as indicated by a color change from red to green. This performance task has been used in other studies to document students’ ability to use evidence and to justify their choice of evidence with an appropriate rationale (Walker et al., 2010; Walker, Sampson, Grooms, Zimmerman et al., in press). In this study, each student analyzed a pair of flasks rather than a set of five flasks. The students emptied the sodium bicarbonate into the flasks, swirled the flasks and observed the balloon inflation and color change. The students did not discuss the demonstration but made independent observations and conclusions. The answer sheet for the performance task (see Appendix B) provided general information about the reaction and asked the students to identify the limiting reactant in each flask, to provide evidence for their conclusion and to justify the use of this evidence with appropriate rationale. Although, this same performance task was used three times, the amount of sodium bicarbonate was varied so that the students evaluated a slightly different scenario with each assessment. The three variations for the performance task are described in Table 4 and shown in Figure 3.

The first assessment was administered during the first laboratory session to establish a baseline for the assessment. The second assessment was administered during the sixth laboratory session prior to the ADI investigation pertaining to mole ratios. It was confirmed that both lecture instructors had completed their coverage of this topic at the time of the second assessment. The third assessment was given during the final lab session.
Table 4

**Description of Performance-based Assessments**

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Flask 1</th>
<th>Flask 2</th>
<th>Description of Reactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2:1</td>
<td>1:2</td>
<td>Flask 1 had excess acid, so the indicator remained red. Flask 2 had excess sodium bicarbonate that was visible as unreacted solid and the indicator changed to green. The balloons were different sizes with the balloon for flask 1 being smaller.</td>
</tr>
<tr>
<td>2</td>
<td>3:2</td>
<td>2:3</td>
<td>Flask 1 had excess acid, so the indicator remained red. Flask 2 had excess sodium bicarbonate that was visible as unreacted solid and the indicator changed to green. The balloons were different sizes with the balloon for flask 1 being smaller.</td>
</tr>
<tr>
<td>3</td>
<td>1:1</td>
<td>2:3</td>
<td>Flask 1 had equal molar amounts of acetic acid and sodium bicarbonate. The indicator turned orange. Flask 2 had excess sodium bicarbonate that was visible as unreacted solid and the indicator changed to green. The balloons were close to the same size.</td>
</tr>
</tbody>
</table>

*Figure 3.* Three variations of the performance-based assessment.
Assessment of Scientific Argumentation in the Classroom (ASAC)

In order to assess argumentation, most researchers video or audio record students as they engage in argumentation, then transcribe the discourse, and finally code or score the transcription using a framework such as Toulmin’s Argument Pattern (Erduran et al., 2004). This process can be difficult for researchers because argumentation is often nonlinear in nature and the various aspects of a verbal argument (e.g. data, warrants, backings) are difficult to identify. Frameworks, such as Toulmin’s Argument Pattern, also tend to ignore social interactions during an episode of argumentation. In consideration of these issues, the video recorded argument generation and evaluation sessions were scored directly using an observation protocol, *Assessment of Scientific Argumentation in the Classroom* (ASAC; see Appendix C) to assess changes in the ways students engaged in argumentation over time. This instrument was designed to capture argumentation events in a more holistic fashion allowing for a more comprehensive assessment of the quality of an argumentation event.

The ASAC is divided into three sections, conceptual and cognitive, epistemic, and social. These sections are based on the integrated domains that Duschl (2008) describes as essential for generating and evaluating arguments in educational contexts:

1. The conceptual structures and cognitive processes used when reasoning scientifically;
2. the epistemic frameworks used when developing and evaluating scientific knowledge; and,
3. the social processes and contexts that shape how knowledge is communicated, represented, argued, and debated (Duschl, 2008, p. 277).

Sampson, Enderle and Walker (accepted with revisions) generated a collection of stem sentences for each of the three domains based on the scientific argumentation literature (e.g. Berland & Reiser, 2009; Driver et al., 2000; Duschl, 2008; Sampson & Clark, 2006). The literature further provided the basis for construction of detailed descriptions for each stem, which supply insight into the characteristics a particular item is intended to measure.

The *Conceptual and Cognitive Aspects of Argumentation* section consists of seven items, which enable the researcher to evaluate the participants’ focus on problem solving, the participants’ evaluation of alternative claims, the participants’ willingness to consider anomalous data, the participants’ skepticism of ideas, the participants’ ability to provide reasons in support of ideas, the participants’ evaluation of alternative claims and in contrast the participants’ use of
inappropriate reasoning strategies. For example, Item 2 in this section states, “The participants sought out and discussed alternative claims or explanations.” This item is designed in part to meet Driver, Newton & Osborne’s (2000) stipulation that individuals need to consider alternative positions when engaged in the construction of an argument. The other items have a similar foundation in research. Scores on this aspect of argumentation can range from 0 to 21.

The Epistemic Aspects of Argumentation section contains seven items, which address the participants’ use of evidence to make sense of the phenomenon, the participants’ evaluation of the evidence, the participants’ use of scientific theories, laws or models, the participants’ ability to distinguish inferences and observations, the participants’ use of the language of science as well as the participants’ use of inappropriate rhetoric to manipulate a claim. In this section there are several items that address the use of evidence. This emphasis is supported in part by Berland and Reiser’s (2009) assertion that if students do not develop a reliance on evidence then they are likely to reaffirm their misconceptions. Scores on this aspect of argumentation can range from 0 to 21.

The Social Aspects of Argumentation section contains six items, with which the researcher will be able to evaluate the interactions of the participants. These items assess the participants’ ability to be reflective about what they say, their respect for each other, their willingness to discuss ideas introduced by others, their willingness to solicit ideas from others and their overall participation. This section as a whole focuses on the participants’ accountability to each other rather than an authority such as the teacher (Kuhn & Reiser, 2006). This section provides a means for assessing the ways participants communicate with each other as they come to consensus on their argument. Scores on this aspect of argumentation can range from 0 to 18.

Two episodes of argumentation were video recorded during each ADI investigation and later evaluated using the ASAC. The first episode occurred as the core group finalized their claim with supporting evidence in the construction of their poster. This argument generation (step 3) required students to collaborate and come to consensus regarding the use of their evidence in supporting a claim. This was followed by an argumentation session (step 4) that involved splitting of the core group into the traveling members and one presenting member. The traveling members typically visited three tables resulting in formation of a transitory group with the presenter at that table. This new group of students engaged in the second type of argumentation that was more persuasive and evaluative in nature, before the travelers moved to
the next table. Following the argumentation session the core group reconvened and discussed their results in light of their discussion with the other groups. This sometimes resulted in a claim change or a reinterpretation of evidence. This discussion also was recorded and scored as part of the argument generation. The argumentation session with the transitory groups was scored as a single group of participants engaged in the presentation and evaluation of an argument.

**Written Arguments**

Writing to learn has been presented as a valued tenet of the ADI instructional model (Sampson et al., in review), however, this research looked specifically at the role of writing in the development of argumentation as practiced in science. While the group argument generation and evaluation provided opportunity for evaluation of evidence and revision of claim, the written argument requires the student to commit to a claim that they then support with genuine evidence and appropriate rationale. In addition, evaluating the written arguments of students provided an assessment of individual argumentation skill, although the influence of collaboration outside the laboratory and the peer review prevented the written argument from being considered wholly individual.

The laboratory reports consist of three sections, section one describes the problem and the context, section two the methodology and section 3 is the argument. Only section three of the laboratory report was relevant for this study. Section three required the student to provide and support an accurate explanation with appropriate evidence and reasoning (i.e., the argument or the product of the inquiry) to the guiding question. The scoring rubric (see Appendix D) for section three examined the following five aspects of the student’s written argument.

1. The author provided a well-articulated claim that provides an adequate and accurate answer to the research question.
2. The author uses genuine evidence to support the explanation and presents the evidence in an appropriate manner
3. The author provides enough evidence to support the claim AND the evidence is valid and reliable.
4. The author’s rationale is sufficient and appropriate.
5. The author compared findings with other groups in their lab section.
As with the ASAC, these five aspects were grounded in the argumentation literature, specifically the epistemic aspects of argumentation as demonstrated by the emphasis on the use of and evaluation of evidence (Driver et al., 2000).

**Data Analysis**

Quantitative measures were used to evaluate each type of data. The performance task and laboratory reports were scored with an established rubric and the video recordings were scored using the ASAC observation protocol. An independent researcher was used to validate the scoring for each data source and the inter-rater reliability of the ASAC and the other two rubrics was assessed using Cohen’s Kappa. Cohen's Kappa is a statistical measure of inter-rater agreement and is generally thought to be a more robust measure than simple percent agreement calculation because it takes into account the agreement occurring by chance. Cohen’s Kappa values range from 0 to 1 and a value of greater than 0.6 indicates good to excellent agreement between two raters (Landis & Koch, 1977).

During the semester students withdrew from the lab, were absent and failed to turn in assignments making the actual numbers collected for each assessment vary to some extent. In addition, over the course of the semester Section A lost five students that resulted in two of the groups being reduced to two members. A third group in section A had intermittent attendance by two members, so three groups from this section were not scored for argument generation or evaluation. Section B retained enough students for all six groups to be evaluated for the entire semester.

**Performance-based Assessment**

The scoring rubric for the performance task (see Appendix B) awarded a single point for correctly identifying the limiting reactant in each balloon, three points for evidence (color, balloon size, and solid residue) and three points for providing a valid rationale for each piece of evidence. The total score possible on this task was seven points for each flask or 14 points per assessment. The performance task was given to the students three times during the semester. All of the answer sheets were scored by the researcher. A second scorer, a science educator who did not have a stake in this study scored 30% of the answer sheets and a Cohen’s Kappa value of 0.69 was obtained indicating good to excellent agreement.

Figure 4 is an example the scoring of student answer for the yellow balloon that had the 2:1 ratio of acetic acid to sodium bicarbonate. The student correctly identified sodium
bicarbonate as the limiting reactant and provided all three pieces of evidence; however their rationale failed to fully explain why this evidence supported their claim.

<table>
<thead>
<tr>
<th>Balloon Color: <strong>Yellow</strong></th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the Limiting Reactant H(_2)C(_2)O(_2) or NaHCO(_3) or Neither? <strong>NaHCO(_3)</strong></td>
<td>1/1</td>
</tr>
<tr>
<td>How do you know? Defend your explanation with appropriate evidence and a rationale.</td>
<td></td>
</tr>
<tr>
<td>The evidence is…</td>
<td>3/3</td>
</tr>
<tr>
<td>The water didn’t change colors and produced a small amount of gas, no substance was visible after the reaction.</td>
<td></td>
</tr>
<tr>
<td>The rationale is…</td>
<td>1/3</td>
</tr>
<tr>
<td>There wasn’t enough NaHCO(_3) to completely react with H(_2)C(_2)O(_2).</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>5/7</strong></td>
</tr>
</tbody>
</table>

*Figure 4. Answer for one flask in the mid-term performance-based assessment.*

**Assessment of Scientific Argumentation in the Classroom (ASAC)**

The video recordings for the five argument generation and argument evaluation sessions were evaluated separately using the ASAC observation protocol. The researcher took notes while watching each video recorded event. Immediately after watching the video each item on the ASAC was scored for that event using a 4-point scale based on how often a particular aspect of argumentation was observed (0 = not observed to 3 = observed often). The ASAC has two items for assessment of inappropriate aspects of argumentation, so for these the scale is reversed (3 = not observed to 0 = observed often). Both events for nine groups were scored by the researcher for a total of 90 overall argumentation scores and 180 domain scores. A fellow researcher who participated in development of the ASAC but who did not have a stake in this study scored both events for one group, representing 11% of the data. Comparison of the scores on each item yielded a Cohen’s Kappa value of 0.60 indicating good to excellent agreement.

In order to determine individual ASAC scores from the argument evaluation data, each table was scored based on the argumentation that took place between the presenter and the different travelers. The score for the table was given to the presenter and the average of the table scores for the tables visited by the travelers was given to each traveler. An example of this
system is presented in Tables 5 and 6. The presenter from table one shared his group’s argument with the travelers from tables 4, 5, and 6 and therefore received a composite ASAC score based on the discussion with all three transient groups (Table 5). The travelers from table one visited tables two, three and six, so their ASAC score was an average of the scores from these three tables (Table 6).

Table 5

*Composite ASAC Scores for the Density Investigation for the Presenting Students*

<table>
<thead>
<tr>
<th>Table</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presenter’s Score</td>
<td>22</td>
<td>22</td>
<td>26</td>
<td>32</td>
<td>26</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 6

*Generation of Individual ASAC Scores for Density Investigation for the Traveling Students*

<table>
<thead>
<tr>
<th>Students</th>
<th>Tables Visited and Score</th>
<th>Average Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1 Travelers</td>
<td>- 22 26 - - 19 22</td>
<td>22</td>
</tr>
<tr>
<td>Table 2 Travelers</td>
<td>- - 32 26 19 26</td>
<td>26</td>
</tr>
<tr>
<td>Table 3 Travelers</td>
<td>- - 32 26 19 26</td>
<td>26</td>
</tr>
<tr>
<td>Table 4 Travelers</td>
<td>22 22 26 - - - 23</td>
<td></td>
</tr>
<tr>
<td>Table 5 Travelers</td>
<td>22 22 26 - - - 23</td>
<td></td>
</tr>
<tr>
<td>Table 6 Travelers</td>
<td>22 - - 32 26 - 26</td>
<td></td>
</tr>
</tbody>
</table>

**Written Arguments**

Finally, a rubric to score the overall quality of the laboratory reports has been described by Walker et al. (in press). As discussed above, in this study only the argument portion of the report was scored. The rubric provides a basis for scoring components of the argument on a scale of 0 (not observed) to 3 (indicating all criteria were met), making a total score of 0 to 15 possible (see Appendix D). A colleague at the community college who teaches the general chemistry laboratories collaborated in scoring five reports from each ADI laboratory, 14% of the reports. A comparison of the scores on each item yielded a Cohen’s Kappa value of 0.83 indicating good to excellent agreement. Figure 5 is the argument section of a student laboratory report. The capital words in brackets precede the relevant rubric criteria.


<table>
<thead>
<tr>
<th>[DATA]</th>
<th>Original Mass</th>
<th>Burn 1</th>
<th>Burn 2</th>
<th>Burn 3</th>
<th>% Water</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 1</td>
<td>55.630 g</td>
<td>55.425 g</td>
<td>55.399 g</td>
<td>55.335 g</td>
<td>34.46%</td>
</tr>
<tr>
<td>Test 2</td>
<td>57.761 g</td>
<td>57.635 g</td>
<td>57.508 g</td>
<td>57.440 g</td>
<td>33.83%</td>
</tr>
<tr>
<td>Test 3</td>
<td>55.681 g</td>
<td>55.432 g</td>
<td>55.396 g</td>
<td>55.387 g</td>
<td>35.13%</td>
</tr>
</tbody>
</table>

[CLAIM] Our average value for percent of water in the hydrate was 34.47%. Because the percent water in copper (II) sulfate pentahydrate was 36.09%, we identified our unknown as copper (II) sulfate pentahydrate. [EVIDENCE] We determined this because our values for percent of water were very close to that of copper (II) pentahydrate. [RATIONALE] Our average value for the percent water was lower than the actual percent water. Our group concluded that if we had taken more time to heat the material, more water would have been allowed to evaporate, thus giving us a more accurate anhydrous mass; furthermore, the percent error would have been less than we calculated. For the error for each trial we calculated, 5.56%, 7.29% and 3.73%, respectively. [COMMUNITY] The other group that received the same hydrate as us also concluded that it was copper (II) sulfate pentahydrate. Because their values and percent error were near ours, it verified our calculations were correct.

Total 12/15

Figure 5. Example of written argument scoring.

The student lost points on his presentation of data, because he did not provide the actual mass values for anhydrous salt, hydrate and water. All of his values were total masses including the test tube. The student lost a single point for his comparison with the scientific community because he did not provide details to support his assertion that the values were “close”. Also, the inference that agreement is indicative of correctness is flawed.

As time was the independent variable in this study a repeated measure of analysis of variance was used to evaluate the significance of student improvement in each area (performance task, argumentation, and written argument) over the course of the semester (Trochim, 1999). In addition, a linear regression analysis was conducted to evaluate how well the ASAC scores predicted individual scores on the written arguments (Green & Salkind, 2005).
CHAPTER 3
RESULTS

Introduction

This chapter presents the statistical analysis for each data set (e.g., performance-based assessment, written argument and ASAC) along with the appropriate tables and figures. As this was longitudinal study a repeated measures ANOVA was conducted for each data set. The preset alpha level of .05 was used to determine statistical significance except with follow-up contrasts. The significance of the follow-up contrasts was determined using the Bonferroni approach of dividing the \( p \)-value by the number of contrasts in order to control for Type I error. The graphical presentation of the mean scores is shown with bars representing ± 1 SD (or 68% of the scores) in order to illustrate the amount of variation in student performance.

Quality of the Arguments Generated on the Performance-based Assessment

A repeated measures ANOVA was conducted for the performance task with the factor being time of the assessment and the dependent variable being the total score. The means and standard deviations for the performance task scores are presented in Table 7. The results of the ANOVA indicated a significant time effect, Wilks’s \( \Lambda = .67 \), \( F(2,29) = 7.06 \), \( p < .01 \), multivariate \( \eta^2 = .33 \). The results of the follow-up contrasts using a series of paired-samples \( t \) tests indicated that the mean score for the mid-term, \( t(33) = 3.59 \), \( p < .01 \) and post-test assessment, \( t(32) = 3.38 \), \( p < .01 \), were significantly different than the pre-test. There was not a significant difference between the mid-term and post-test assessments. Using the Bonferroni approach to control for Type I error across the three contrasts, a \( p \) value of less than .02 (\( p = .05/3 = .02 \)) was required for significance. The means and standard deviations are presented graphically in Figure 6.

Table 7
Means and Standard Deviations for Performance-based Assessment Scores

<table>
<thead>
<tr>
<th>Event</th>
<th>( N )</th>
<th>Range</th>
<th>( M )</th>
<th>( SD )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Test</td>
<td>36</td>
<td>0 – 9</td>
<td>3.64</td>
<td>1.92</td>
</tr>
<tr>
<td>Mid-Term</td>
<td>35</td>
<td>1 – 9</td>
<td>5.03</td>
<td>2.02</td>
</tr>
<tr>
<td>Post-Test</td>
<td>34</td>
<td>1 – 12</td>
<td>5.50</td>
<td>2.90</td>
</tr>
</tbody>
</table>
A repeated measures ANOVA was conducted for each of the components of the performance task with the factor being time of the assessment and the dependent variable being the answer, evidence or rationale. The results of the ANOVA indicated a significant time effect, Wilks’s $\Lambda = .45$, $F(2,29) = 17.69$, $p < .01$, multivariate $\eta^2 = .55$ for the answer and a significant time effect, Wilk’s $\Lambda = .71$, $F(2,29) = 5.87$, $p < .01$, multivariate $\eta^2 = .28$ for the rationale. The ANOVA for the evidence component was not significant. The means and standard deviations for the component scores are presented in Table 8.

Table 8

<table>
<thead>
<tr>
<th>Event</th>
<th>$N$</th>
<th>Answer Mean</th>
<th>SD</th>
<th>Evidence Mean</th>
<th>SD</th>
<th>Rationale Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Test</td>
<td>36</td>
<td>.81</td>
<td>.749</td>
<td>1.92</td>
<td>1.32</td>
<td>.92</td>
<td>.94</td>
</tr>
<tr>
<td>Mid-Term</td>
<td>35</td>
<td>1.66</td>
<td>.639</td>
<td>2.26</td>
<td>1.12</td>
<td>1.11</td>
<td>1.05</td>
</tr>
<tr>
<td>Post-Test</td>
<td>34</td>
<td>1.44</td>
<td>.746</td>
<td>2.38</td>
<td>1.37</td>
<td>1.68</td>
<td>1.27</td>
</tr>
</tbody>
</table>
The results of the follow-up contrasts using a series of paired-samples $t$ tests are presented in Table 9. The means and standard deviations for each component are presented graphically in Figure 7. Using the Bonferroni approach to control for Type I error across the six contrasts, a $p$ value of less than .01 ($0.05 / 6 = .008$) was required for significance. The pre-test mean score for the answer was only significantly different from the mid-term mean. The mean rationale score was only significantly different for the pre-test vs. post-test comparison.

Table 9

*Results of Paired Samples $t$-Tests for Performance-based Assessment Component Scores*

<table>
<thead>
<tr>
<th>Component</th>
<th>Event</th>
<th>Mean Difference</th>
<th>$t$</th>
<th>$p$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answer</td>
<td>Pre-test</td>
<td>.85</td>
<td>6.35</td>
<td>&lt;.001</td>
<td>1.1</td>
</tr>
<tr>
<td>Answer</td>
<td>Mid-term</td>
<td>.61</td>
<td>3.12</td>
<td>.04</td>
<td>.54</td>
</tr>
<tr>
<td>Answer</td>
<td>Post-test</td>
<td>-.28</td>
<td>-1.72</td>
<td>.10</td>
<td>NA</td>
</tr>
<tr>
<td>Rationale</td>
<td>Pre-test</td>
<td>.21</td>
<td>1.05</td>
<td>.30</td>
<td>NA</td>
</tr>
<tr>
<td>Rationale</td>
<td>Mid-term</td>
<td>.70</td>
<td>3.17</td>
<td>&lt;.001</td>
<td>.55</td>
</tr>
<tr>
<td>Rationale</td>
<td>Post-test</td>
<td>.53</td>
<td>2.52</td>
<td>.02</td>
<td>.45</td>
</tr>
</tbody>
</table>

**Quality of the Written Arguments Produced in Laboratory Reports**

A repeated measures ANOVA was conducted for the argument section of the laboratory reports with the factor being time and the dependent variable being the argument score. The sequence of laboratory investigations provided five events over the 10 weeks of data collection. The means and standard deviations for the written argument scores are presented in Table 10. The results of the ANOVA indicated a significant main effect for time, Wilks’s $\Lambda = .46$, $F(4,31) = 9.05, p < .01$, multivariate $\eta^2 = .54$. The means for each investigation $\pm 1$ SD are presented in Figure 8. The results of this analysis indicate the quality of the arguments included in the investigation reports improved over time.
Figure 7. Scores on each component of the performance task.

Table 10

Means and Standard Deviations for Written Arguments

<table>
<thead>
<tr>
<th>Investigation</th>
<th>N</th>
<th>Range</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density</td>
<td>35</td>
<td>1–15</td>
<td>7.74</td>
<td>3.58</td>
</tr>
<tr>
<td>Hydrate</td>
<td>34</td>
<td>2–14</td>
<td>6.85</td>
<td>3.29</td>
</tr>
<tr>
<td>Solutions</td>
<td>32</td>
<td>1–15</td>
<td>7.91</td>
<td>4.15</td>
</tr>
<tr>
<td>Mole Ratio</td>
<td>33</td>
<td>2–15</td>
<td>9.12</td>
<td>4.41</td>
</tr>
<tr>
<td>Reactions</td>
<td>29</td>
<td>3–15</td>
<td>9.52</td>
<td>4.03</td>
</tr>
</tbody>
</table>
Quality of the Argumentation Episodes That Took Place During the Investigations.

A repeated measures ANOVA was conducted for the two argumentation events that take place during ADI instructional model with the factor being time and the dependent variable being the ASAC score. Again, the five ADI investigations that took place over the course of the semester provided a series of events to evaluate for change over time. The means and standard deviations for the ASAC scores are presented in Table 11. The results of the ANOVA indicated a significant main effect of time for the argument generation events, Wilks’s Λ = .13, F(4,5) = 8.379, p = .02, multivariate η² = .87, as well as the argument evaluation events, Wilks’s Λ = .10, F(4,5) = 10.79, p = .02, multivariate η² = .90. As illustrated in Figure 9, the ASAC scores for both argument generation and argument evaluation follow a similar trend that is higher for the first investigation than the second then increases for the third and fourth investigations with slight drop for the final investigation. The means and standard deviations for the cognitive, epistemic and social sections of the ASAC scores are presented in the Table 12.
Table 11

Means and Standard Deviations for ASAC Scores (N = 9)

<table>
<thead>
<tr>
<th>Investigation</th>
<th>Range</th>
<th>M</th>
<th>SD</th>
<th>Range</th>
<th>M</th>
<th>SD</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density</td>
<td>16 – 32</td>
<td>25.78</td>
<td>5.07</td>
<td>8 – 30</td>
<td>22.67</td>
<td>6.34</td>
<td>3.11</td>
</tr>
<tr>
<td>Hydrate</td>
<td>13 – 27</td>
<td>19.44</td>
<td>4.88</td>
<td>11 – 25</td>
<td>17.78</td>
<td>5.12</td>
<td>1.66</td>
</tr>
<tr>
<td>Solutions</td>
<td>14 – 35</td>
<td>24.82</td>
<td>7.05</td>
<td>11 – 37</td>
<td>24.62</td>
<td>7.86</td>
<td>0.20</td>
</tr>
<tr>
<td>Reactions</td>
<td>17 – 27</td>
<td>21.76</td>
<td>2.91</td>
<td>19 – 32</td>
<td>25.26</td>
<td>4.66</td>
<td>-3.50</td>
</tr>
</tbody>
</table>

Figure 9. ASAC scores for both argumentation events.
Table 12

Means and Standard Deviations Component ASAC Scores (N = 9)

<table>
<thead>
<tr>
<th>Investigation</th>
<th>Cognitive</th>
<th>Epistemic</th>
<th>Social</th>
<th>Cognitive</th>
<th>Epistemic</th>
<th>Social</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M  SD  M SD</td>
<td>M  SD</td>
<td>M SD</td>
<td>M  SD  M SD</td>
<td>M  SD</td>
<td>M SD</td>
</tr>
<tr>
<td>Density</td>
<td>7.13  1.73</td>
<td>11.63  2.39</td>
<td>7.25  2.66</td>
<td>6.78  2.39</td>
<td>10.78  2.73</td>
<td>5.11  1.96</td>
</tr>
<tr>
<td>Hydrates</td>
<td>6.56  2.60</td>
<td>8.89  1.45</td>
<td>4.44  2.07</td>
<td>5.33  1.41</td>
<td>7.22  3.19</td>
<td>5.22  1.64</td>
</tr>
<tr>
<td>Solutions</td>
<td>7.63  3.07</td>
<td>11.38  2.56</td>
<td>5.38  3.81</td>
<td>7.38  2.72</td>
<td>10.75  4.10</td>
<td>6.50  1.93</td>
</tr>
<tr>
<td>Mole Ratio</td>
<td>9.29  2.81</td>
<td>12.57  4.43</td>
<td>6.86  3.34</td>
<td>8.88  2.36</td>
<td>12.63  2.50</td>
<td>9.25  3.11</td>
</tr>
<tr>
<td>Reactions</td>
<td>6.78  1.58</td>
<td>10.25  1.16</td>
<td>1.78  3.11</td>
<td>7.38  1.77</td>
<td>11.00  1.85</td>
<td>6.88  1.96</td>
</tr>
</tbody>
</table>

The Relationship Between the Argumentation Events and the Written Arguments

Correlation coefficients were first calculated between the Written Argument, Argument Evaluation and Argument Generation scores in order to determine the extent of collinearity between these three variables. Using the Bonferroni approach to control for Type I error across the three correlations, a p value of less than .02 (.05 / 3 = .02) was required for significance. The results of the correlational analyses presented in Table 13 show that all three correlations were statistically significant and greater than or equal to .30. The significant correlation coefficients suggest that the argumentation event variables are collinear, which violates the independence assumption for a multiple regression analyses. The Argument Evaluation score had a slightly stronger correlation with the written argument score and was therefore used for the regression analysis.

A linear regression analysis was then conducted to determine how well the ASAC scores for Argument Evaluation events predict the written argument score on the investigation reports. The regression equation for predicting the written argument score is:

\[ \text{Written Argument Score} = .23 \text{ ASAC Score} + 3.18 \]

The 95% confidence interval for the slope, .10 to .35 does not contain the value of zero, and therefore the written argument score is significantly related to the Argument Evaluation ASAC

37
Table 13

*Correlations among the Three Argument Scores (N = 95)*

<table>
<thead>
<tr>
<th></th>
<th>Written Argument</th>
<th>Argument Generation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argument Generation</td>
<td>0.32*</td>
<td></td>
</tr>
<tr>
<td>Argument Evaluation</td>
<td>0.35*</td>
<td>0.30*</td>
</tr>
</tbody>
</table>

* p < .02

Accuracy in predicting the written argument score was moderate. The correlation between the Argument Evaluation ASAC score and the written argument score was 0.35. The scatter plot for the two variables, as shown in Figure 10, indicates that the two variables are linearly related such that as argument evaluation ASAC score increases so does the written argument score. Approximately, 12% of the variance in written argument was accounted for by the linear relationship with the Argument Evaluation ASAC score.

*Figure 10. Scatter plot for argument evaluation scores vs. written argument scores.*
CHAPTER 4
DISCUSSION

This chapter provides a discussion of the results I presented in Chapter Three. The discussion of the results is organized around each research question.

**Question 1 – Performance-based Assessment**

*To what extent does the ADI instructional model improve undergraduate student's performance on a task that requires them to develop and support a valid conclusion with appropriate evidence and an adequate rationale?*

There was a significant positive linear trend over the course of the semester, indicating that student scores on the performance task did improve. This finding suggests that the ADI instructional model does facilitate individual growth in generation of an argument. Pairwise comparisons between the mid-term and post-test assessment showed them to be significantly different from the pre-test, but not from each other. There are several reasons that the improvement between the second and third assessment might not be significant. First, and foremost, the third assessment was the most challenging of the three. The difference between the two flasks was less obvious, the balloons inflated to roughly the same size and one flask did not remain red, but turned orange. The instructions told the students that red indicated acid was present and green indicated that no acid was present. The students had to infer that orange indicated a stoichiometric mix of the reactants, such that neither was limiting. Their difficulty with this is reflected in the slight drop in the average answer score for the final assessment. The rationale score, in contrast, showed improvement from the midterm assessment to the post-test. This is similar to the findings of other research that has shown that students have difficulty providing the backing, or what we refer to as rationale, for why they chose to use certain pieces of data as evidence in their written arguments (Bell & Linn, 2000; McNeill, Lizotte, Krajcik, & Marx, 2006). McNeill & Krajcik (2007), for example, reported that the middle school students in their study learned how to support their claim with evidence quickly, but that they less frequently articulated the scientific principles that allowed them to make that connection. The numerous studies on scientific explanation and argumentation have suggested that this ability does not come naturally to most individuals but rather is acquired through practice (Osborne et al., 2004). As with the Nature of Science, educators debate the need for students to be explicitly taught how to construct an argument complete with claim, evidence and rationale or whether
instructional methods can be used to foster these skills over time (Jimenez-Aleixandre, 2007). The results of this study would suggest that the latter case has potential for developing argumentation skills. The observation that students develop their argument skill in the same, well documented, pattern of claim with evidence first with later development of the rationale lends further support to the use of the ADI instructional model for engendering student understanding and use of this key scientific process.

Student knowledge and understanding of chemical reactions could provide an additional or complementary explanation for the delayed development of the rationale in student arguments on the performance task. Metz (2000) points out that the ability to construct an accurate scientific explanation represents a complex interaction of scientific reasoning capacities and domain-specific knowledge. The second performance-based assessment was given after the concept of limiting reactant had been covered in the lecture course, but prior to the ADI investigation on the topic. It seems reasonable to suggest that the ADI investigation augmented the domain-specific knowledge to such a degree that when coupled with continuous experience with argumentation provided in the laboratory that the students provided a better rationale for the claim on the final performance task.

In all three iterations of the performance task one of the balloons contained excess sodium bicarbonate. This repeated scenario provided a consistent point for comparison of student understanding of the content and their skill at constructing an argument. This progression is illustrated here by the responses of a single student whose total scores on the three performance tasks reflected the mean for each assessment. This student provided the following answers the first time she completed the performance task:

*Is the Limiting Reactant HC₂H₃O₂ or NaHCO₃ or Neither?* Neither

*The evidence is...* The concentrate of HC₂H₃O₂ changed to a yellow color. There may have been minimal contents of acid in the NaHCO₃, enough to chemically change the color to yellow, but not enough to turn green

*The rationale is...* Because no acid concentrate would show by no color change in HC₂H₃O₂, and the contents in the vile did change color, there had to have been some other chemical – if not acid- that reacted with the HC₂H₃O₂

On this task the student earned a total of one point of the seven possible points. First, she
incorrectly identified neither reactant as limiting (0/1 point). As evidence, she only lists the yellow color (1/3 points). Finally, she proposes a third mystery chemical that prevented the reaction from turning completely green for her rationale instead of attempting to justify her claim with the evidence (0/3 points). Her answers clearly represent a sincere attempt to complete the performance task but also illustrate her limited understanding of the concept and the components of an argument. The answer is incorrect, the evidence is minimal and the rationale is illogical.

At the midpoint in the semester, she provided the following answers on the performance task:

*Is the Limiting Reactant HC$_2$H$_3$O$_2$ or NaHCO$_3$ or Neither?*  HC$_2$H$_3$O$_2$

*The evidence is...* The acid turned green.

*The rationale is...* The base reacted with the acid, and because there was an excess amount of base, the solution turned green.

She correctly identified acetic acid as the limiting reactant (1/1 point), but she again simply used an observed color change as evidence (1/3 points). Although she only included a single piece of evidence to support her claim, she no longer uses an unsubstantiated inference as evidence, which is an improvement over her first attempt. Her rationale is sound but minimal (1/3 points) giving her a total score of three out of seven points possible. This would then be a threshold where the student has the content knowledge, but not the depth of understanding or the skill to construct a robust argument.

For the third performance task, which was the most challenging of the three, she provided the following answers:

*Is the Limiting Reactant HC$_2$H$_3$O$_2$ or NaHCO$_3$ or Neither?*  HC$_2$H$_3$O$_2$

*The evidence is...* There was a much greater amount of NaHCO$_3$ in this reaction. Because the formula provided is balanced and is an optimum mole ratio, having more NaHCO$_3$ made HC$_2$H$_3$O$_2$

*The rationale is...* Green indicates that no acid is present because the entirety of HC$_2$H$_3$O$_2$ was allowed to completely react. There was more NaHCO$_3$ in the balloon than in the orange balloon. There was still some NaHCO$_3$ that was not dissolved.

On the final performance task this student constructed a robust argument in support of a
correct answer (1/1 point). She only identifies excess sodium bicarbonate specifically as evidence (1/3 point), however she uses all of the possible evidence in the rationale (3/3 points). The evidence score on the performance task was assessed using only the information written directly under evidence, on occasion as in this case additional evidence was used in the rationale that was not used in the evidence score. Her total score for this flask was five out of the seven points possible. Anecdotally, both the researcher and the second scorer noted the improvement in the student rationales for the third performance task.

These results support the notion that a well designed integrated instructional model, such as ADI, can impact an individual student’s ability develop and support a valid conclusion with appropriate evidence and an adequate rationale. In addition, as noted by others (Berland & Reiser, 2009; Sandoval & Millwood, 2005) the ability to provide a sound backing or rationale is slower to develop for most students. Science educators should keep in mind that this key aspect of argumentation does not come easily, but develops over time with continual exposure and guidance.

**Question 2 – Written Arguments**

*To what extent does the ADI instructional model improve the quality of the scientific arguments undergraduate students write over the course of a semester?*

There was a significant positive linear trend in written argument scores over the course of the semester, indicating that the ADI instructional model did facilitate improvement in the quality of the arguments the students included in their investigation reports. To illustrate the improvement in the students’ arguments over time, consider the following series of arguments that were written by the same student over the course of the semester. This student was chosen because his scores closely followed the class trend. This student consistently scored the maximum three points on the data presentation component of the argument, so only the text of their argument is given. The capital words in brackets are placed just before the relevant rubric criteria. Table 14 provides a summary of the student’s scores for each investigation.

*Density Example*: Overall, the participants in this study appeared to understand the importance of presenting their data in tables or graphs so that as in this example most scored consistently high on this aspect of the rubric. Most students also included a claim or conclusion in their argument. Many of these students, however, did not provide a sound rationale for their
### Table 14

**Written Argument Scores for a Single Student**

<table>
<thead>
<tr>
<th>Investigation</th>
<th>Claim</th>
<th>Data</th>
<th>Evidence</th>
<th>Rationale</th>
<th>Community</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Hydrate</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Solutions</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Mole Ratio</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>Reactions</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>14</td>
</tr>
</tbody>
</table>

To illustrate this trend, consider the following excerpt. This excerpt was typical of arguments that were included in the first report. The guiding question for this investigation asked, “Are these objects made of the same material?”

At the conclusion of this lab investigation, we evaluated the results of our data. [RATIONALE] Based on close similarities in densities, [CLAIM] we hypothesized that both the gray cube and the purple cylinder were made from the same substance, whereas the red cylinder was made from a differing material. [EVIDENCE] By performing two methods of measurement in this lab, we were able to determine the best measuring tool and provide an argument to explain our findings. Our density calculations from both the Vernier caliper and water displacement experiments were relatively similar. However, the caliper experiment proved to be much more accurate and would be recommended due to its strong validity and reliability. If all the supplied objects sank/had a greater density than water as did the majority of the other groups’ objects, we speculate that the calculations would have been much closer due to the margin of error being reduced.

This student clearly answered the guiding question. Yet, there was little backing to the claim other than that it was based on similarities in densities. The density values were not used in the conclusion, nor was there any evaluation regarding the alleged similarities. From field observations I know that students often assume that the presentation of the data is sufficient and that the reader will refer to the data table as needed. Rather than using the data to support the claim, the student argued the validity of the methodology by making generalized claims as to accuracy and precision. The analysis of reports indicated that many of the participants in this study wrote arguments that were similar, in terms of content and structure, to this example for
the first report. Overall, these students consistently answered the question, but failed to provide a rationale and in this first report they tended to discuss their methods extensively.

**Hydrate Example:** For most students, the second laboratory report did not differ much from the first with regard to the elements of argument included or missing. There was a claim and a data table, but very little backing and little or no comparison with their matched group for the unknown. The second investigation asked the students to identify an unknown hydrate compound as one of four possible compounds by using the percent water.

Our hypothesis was that if the percentages of water in our sample and one of the provided hydrates matched, then they were the same substance. [CLAIM] At the conclusion of the experiment, we determined that our unknown hydrate was BaCl$_2$·2H$_2$O.

This excerpt is remarkably succinct and while there were examples of student arguments with a bit more detail in general this argument was representative. The first sentence sets up the basis for a rationale but there is no follow through. The percentage of water that would be consistent with the claim of barium chloride dehydrate was not included in the argument, nor was the experimental value provided as evidence. As in this excerpt, comparison with other groups was typically not included by students in the second laboratory report.

**Solutions Example:** The third laboratory investigation was a more challenging investigation for the students in that it required them to integrate two techniques that were unfamiliar to most students, spectroscopy and chromatography. The guiding question was, “How could I prepare more of this dye?” In order to answer this question, students needed to identify the components of their solution using paper chromatography and determine the concentration of these components using spectroscopy. Integrating the data from both sources was challenging for most students; an issue that was reflected in their written arguments.

Ultimately, by ascertaining the identity and concentrations of each dye, we were able to identify the unknown purple dye as a mixture of two colors. We determined that the unknown purple dye contained 5.65E-6 M blue dye B1 and 9.60E-6 M red dye R3. [EVIDENCE] By comparing our calculations, the chromatogram illustrated that our solute distances of the blue B1 and red R3 dyes lined up with the unknown purple dye [components]. In order to make more of the unknown purple dye and solve the lab problem, one could take the molarities of each dye and make a solution. [CLAIM] To make 1 liter of solution, one could take 5.65E-6 moles of blue dye B1 and 9.60E-6 moles of red dye R3 and add enough water to make a 1L solution. If more dye is required, the same ratios should be used with increasing amount of solution. [COMMUNITY] In complying with good scientific practice, we compared our results with those of
the other group that had the same unknown. They advised us that they performed the same process and also received the same results. Being that they also determined the unknown purple dye to contain blue dye B1 and red dye R3 our results are supported and we are more confident in our findings.

This excerpt was typical of student arguments for the solutions investigation. As in this case most students focused on one method, spectroscopy or chromatography, but not both. The identification of the B1 and R3 with chromatography was discussed, but the concentration values were dropped into the argument with no discussion or basis. This student does go into detail on actually preparing the solution, in contrast other students simply stated that a solution of the given component concentrations could be prepared. This argument still lacks backing and the evidence is incomplete. However, for the first time the student compared their results with another group and in a relatively robust manner. Typically students made a general statement as to the consistency of results and then inferred accuracy of their claim. This excerpt uses the group comparison in a manner that demonstrates that the student is not looking for correctness but support for a claim. Overall, this excerpt illustrates the transition from the traditional answer on a lab report presented in hydrate example to an actual argument.

*Mole Ratio Example:* This laboratory investigation was simpler in terms of technique and data acquisition; however, the results were more ambiguous. The guiding question asks “What is the optimum mole ratio for a chemical reaction?” The students had to compare the volume of carbon dioxide collected using varying molar amounts of sodium bicarbonate and acetic acid. The answer of course was a 1:1 ratio, but the amount of carbon dioxide gas actually increased slightly when sodium bicarbonate was in excess. The students had to come to grips with the idea that “more” was not necessarily “optimum”.

In comparing our calculations, [CLAIM] we determined the optimum mole ratio to be 1:1 because the smallest number of moles was used to receive the largest amount water displacement from the production of carbon dioxide. [EVIDENCE] For the 1:1 ratio, only 2.10 g of sodium bicarbonate were used to produce 462 mL of carbon dioxide. The 3:1 and 2:1 ratios containing smaller masses of sodium bicarbonate produced much less water displacement from the carbon dioxide. Although the 1:2 and 1:3 mole ratios containing larger masses of sodium bicarbonate produce large amounts of water displacement from the carbon dioxide, they were still smaller than the 1:1 ratio and much less relative to their given masses. [RATIONALE] After the 1:1 ratio, the water displacement from the carbon dioxide products started to decrease. The larger masses of sodium bicarbonate did not give the best ratio because there was an excess amount of the substance leftover in the flask and all of the acetic acid had reacted. This
illustrated that the acetic acid was the limiting reactant in this chemical reaction because it restricted the amount of product formed and was completely consumed prior to the other reactant. [COMMUNITY] In complying with good scientific practice, we compared our results with the other groups. All but one group determined the optimum mole to mole ratio would also be 1:1. This overwhelming corresponding data within the class confirms that our methods and results are sufficient and acceptable.

This excerpt, for the first time, contains all the elements of an argument. The claim is supported by evidence and a sound rationale. The rationale not only justifies the claim but the evidence used for the claim as well. His comparison with the community was used to support the claim and the methods as being “sufficient and acceptable”. Overall, this report was representative of the peak observed in the linear trend for the written arguments.

Reactions Example: This final investigation asked the students to prepare 0.5 g of silver chloride by precipitation. A significant amount of laboratory time was needed for the students to write the balanced equation and calculate the amounts of reactants needed for the reaction. This investigation was intended to integrate multiple concepts from the semester, and apply them to a real world chemical problem. The guiding question was, “Can you prepare 0.5 g of product?” This excerpt was an excellent example of what could have been argued in this experiment, however it was somewhat atypical. As was evidenced by the slight drop in the mean scores for written arguments, most students answered the question with a minimal argument.

At the conclusion of this lab experiment, [CLAIM] we determined that it was indeed possible to make 0.5 g of AgCl within a small percent error margin. In order to create the amount of AgCl, 3.44 mL CaCl₂ needs to combine with 6.98 mL AgNO₃. The limiting reagent of this chemical reaction was AgNO₃ which ultimately determined how much product was formed because the entire amount of this compound was expended. [EVIDENCE] Although our solid measured over the necessary 0.5 g AgCl, we attribute this difference of 5 thousandths of a gram to either still having liquid in the substance or that it had begun to oxidize. Our percent yield was 101% which gave only a 1% error margin. [RATIONALE] Due to time constraints and the authorization of the teaching staff, we only performed two “heat, cool, weigh” sequences. If the precipitant still contained water that possibly could have been evaporated had we performed another series. [COMMUNITY] In complying with good scientific practice, we compared our results with the two other groups that used the same material. We all performed the experiment in the same manner and the other two groups came within 3% error margin, 97.6% and 97.4%. The commonalities between groups and the relative closeness of our product to the desired 0.5 g supports our prediction that 0.5 g of silver chloride can be made.
As in the mole ratio example, this student incorporated all the desired elements of an argument in answering the question. As discussed above, this excerpt is above the mean for this investigation, but demonstrates that the mole ratio investigation was not unique in producing better arguments overall. This student as well as others incorporated the elements of a written argument into his thinking about and understanding of chemistry.

This series of written arguments illustrates the class trend of a relatively high argument score on the first report followed by a lower score on the second report and subsequent growth over the next two reports and in this case a leveling off of the final score. This student did essentially reach a ceiling with the argument score with 14 points out of 15. This example demonstrates that students typically answer the question and present data, but consideration of the validity of the evidence and the rationale behind the answer requires development as does the use of the scientific community. The third report simply answered the question repeatedly without providing evidence or rationale for the concentration values. The fourth report is remarkable in comparison with the third report in the degree to which the student evaluates the evidence and provides rationale for the claim. This seemingly abrupt adapting of the elements of a scientific argument was noted in the oral argumentation as well.

There are two possible explanations for this observation; one is the nature of the task and the second is the instructional model. The mole ratio investigation is elegant in its simplicity and is a sharp contrast to the more complex solutions investigation that required integration of spectroscopy and chromatography. On the other hand, the fourth report was submitted the seventh week of the semester, following 28 hours of ADI based instruction that demanded students engage in argumentation in multiple contexts. If simplicity of the investigation was the determining factor, then the hydrate investigation would have warranted a better argument. The more likely explanation is that immersion of students in a classroom environment that provides instruction in and promotes the principles of argumentation will lead to student growth in written argument. The student scores on the written arguments level off with the fourth and fifth reports. This is most likely due to the quality of the guiding question in the fifth ADI investigation (Can you make 0.5 g of product?) generating little reason for argument. There also, could be a plateau or ceiling in student written argument in this context. This issue will have to be resolved with further research following modification of the guiding question for the fifth investigation. The examples from the written reports demonstrate the process of growth in argument skill as moving
from a simple answer combined with a data table, to robust answer supported with evidence and a sound rationale.

Overall, these results support the conclusion that use of the ADI instructional model improves the quality of the scientific arguments undergraduate students write over the course of a semester. In addition, the slow development and understanding of the use of evidence and rationale in generating a sound argument is revealed in the laboratory reports as it was in the three performance task assessments. Similarly, the use of the community for supporting a claim also required time to develop. These separate components were difficult to tease out given the small sample size in this study, but based on anecdotal assessment from instructors there is likely to be useful insights into student written argument from a larger sample set.

**Question 3 – Argumentation Generation and Evaluation**

*To what extent does the ADI instructional model improve the quality of the scientific argumentation that takes place between undergraduate students over the course of a semester?*

The ASAC scores for the Argument Generation and Argument Evaluation both showed a significant positive linear trend over the semester (Table 11 and Figure 9). Both argumentation events follow a similar pattern with a relatively high score on the first investigation, then dropping sharply for the second followed by an increase for the next two investigations. The final lab investigation resulted in a slight drop in both ASAC scores. The first investigation on density included an unfamiliar piece of equipment, the spill can. This resulted in a significant amount of discussion on how to use the spill can and why the resulting volume tended to be considerably different from the value obtained with calipers. In addition, some groups received an object with a density less than 1.0 g/mL. This presented the groups with a particular challenge when trying to find volume by water displacement. The unusual nature of the spill can and the floating object led to an unexpected degree of discussion about method that incorporated more elements of argumentation than the other investigations. Students struggled with how to conduct the investigation and repeatedly questioned the validity of the data they collected. The spill can method for water displacement did not provide reproducible results either within a group or between groups. The spout on the cans would drip water when slightly disturbed and this made it difficult obtain a consistent initial volume and to know when the “spilling” was complete. The floating object was handled two ways, either the floating was ignored and
explained away as a “source of error” or the object was pushed under water with paper clip which was considered unreliable but necessary.

The laboratory program includes a pre-laboratory experience between investigations. Typically, the peer review requires an hour of laboratory time leaving a second hour for introduction to equipment, techniques or concepts important for the next investigation. The week preceding the density investigation the students did work with different laboratory tools for measuring mass, volume and length; however, the spill can was not included in this preliminary exercise. Density was the only investigation that students were given an unfamiliar tool to incorporate into their investigation.

The drop in ASAC scores for the final investigation, chemical reactions, was probably due to a weak guiding question, “Can you prepare 0.5 g of product?” The intention was for students to try different combinations of reactants in order to achieve higher yields. The time required to dry the product and obtain an initial yield precluded repeating the experiment for most groups, so that they just reported their yield for stoichiometric amounts of the reactants. In addition, these yields tended to be 80 – 90%, which was a reasonably good yield and as a result did not necessitate more of an answer to the question beyond “yes”. Investigations that do not generate much difference in claim do not engender meaningful argumentation.

Closer examination of Figure 9 reveals a shift in the relative ASAC scores. In the first two investigations the argument generation scores are higher than the argument evaluation scores with mean differences of 3.11 and 1.66 for the Density and Hydrate Investigations respectively. For the third investigation, Solutions, the means only differ by 0.20. In the final two investigations the means have reversed, argument evaluation being higher than argument generation with mean differences of 2.40 and 3.50 for the Mole Ratio and Reactions Investigations respectively. The argument generation scores likely suffered from group familiarity. In watching the video recordings, it was clear that group members took on pre-set roles for each investigation. These roles included whiteboard designer, calculator, chemical gatherer and director. In a typical ADI laboratory section students are shuffled for each new investigation to avoid this situation. The argument evaluation session required students to evaluate claims, evidence and reasoning, constructs that students have less experience navigating (Kuhn & Reiser, 2005). This observation supports the notion that the use of cultural tools, such as the whiteboard session, that are embedded in a classroom facilitates growth in student ability
to evaluate what counts as quality evidence in support of a claim. As the students became familiar with these criteria the ASAC scores suggest that their argumentation improved.

The small $N$ value made it inappropriate to perform a repeated measures ANOVA on the section scores for the ASAC; however, some insight can be gained by visual inspection of the means and standard deviations for the cognitive, epistemic and social components of the ASAC scores provided in Table 12. The cognitive and epistemic sections each had seven items to score making direct comparison possible. The epistemic score was higher for all five investigations in both argumentation events. This can be explained most directly by the relatively few instances of consideration of alternative claims or claim changes, for which the cognitive section has three items. During the argument generation stage of the investigation, the students would discuss their data, the acquisition of the data, the calculations involving the data, but generally once they came to a conclusion they did not consider alternatives. The follow-up discussion after the argument evaluation events, where claim assessment would be most likely, was consistently minimal. Students were eager to leave the lab and often sent someone back to their table to make claim changes or edit data during the argumentation session. This particular issue contributed directly to the low occurrence of “claim change” and consequently the lower cognitive scores.

Some understanding into this situation was gleaned from video discussions during the argument evaluation and the following week regarding the laboratory report. During the argument evaluation the students did not know how to address differences in results and data, they operated with the idea that one answer was right and one answer was wrong and, as a result, the students did not attempt to identify a potential source of error in one or both claims. To illustrate this trend, consider the following discussions that took place during the hydrate investigation, between two groups attempting to identify the same unknown. These groups visited each other’s table at the beginning of the argument evaluation session to compare claims. The first discussion occurred at table 2B with travelers from table 6B (P is the presenter for table 2B and T1-T3 are the travelers from table 6B).

P: Ok, so we got completely different things from yours (*she had gone and looked at their whiteboard earlier*). We got zinc sulfate.

T1: Huh?
P: Obviously we were both supposed to identify the unknown hydrate C, we said it was zinc sulfate because our hydrous amount was 0.508 and our anhydrous amount was 0.304 and after calculating we have 59.65% of salt left, subtracting that by one hundred and we 40.35% for trial A. For trial B we got 45.56%. Those are close enough that we took an average, uh – we could have done more trials, but we had to start over because we initially put all in the bottom instead of spreading it.

T1: Yeah, fantastic, yeah. (nods head as if understanding the problem)

P: Our average was 42.955 and our percent error was 2.086, uh, I don’t know if you did anything different in your calculations that would change the percent. Did you take into consideration that it was supposed to spread throughout the entire (uses hand to show spreading up the side of the test tube).

T1: Yeah, yeah, yeah, we got that, did that.

P: Uhm, I don’t know what it could be then. The mass would not have affected it that much. Did y’all start with 0.508 grams?

T1: Yeah, yeah we did that. Let’s see, what did we do (refers to notes).

T2: We did point five.

T3: Point five.

P: How many trials did you do?

T2: About two, three? (looks to T1 for confirmation)

P: Y’all did three?

T1: We did two separate trials.

P: Yeah, we did everything the same; I don’t know why they’re different. I am trying to think of what would offset it.

T1: You used two different tubes, right?

P: Yeah, and we labeled them.

T1: This is really confusing. I just don’t understand how we got different percents. (T1 and T3 have a conversation off camera that sounds like they are discussing the calculations.)
P: It doesn’t really matter as long as you write down what you did. What was your percent error?

T2: It was point something.

T1: Yeah, point seven percent.

P: Well, yours is definitely more accurate than ours.

This discussion illustrates the focus on correctness and the inability to fully investigate sources of error. The only real source of error in this experiment was derived from failure to heat the sample to a constant mass. The zinc sulfate pentahydrate was 42% water, the reason for table 6B’s low value was that they did not remove all of the water. The participants never discussed the likelihood that the sample was not fully dehydrated. The method of heating was discussed which was a legitimate concern, however the discussion regarding the amount of material heated was irrelevant as the percent water was a constant.

The following discussion took place at table 6B. This discussion demonstrated the reliance on the authority of the instructor to resolve the discrepancy in claim (P’ is the presenter for table 6B and T1’-T3’ are the travelers from table 2B).

P’: Our goal was to identify the unknown hydrate. Our claim was that it was copper two sulfide, I mean sulfate actually. The experiment that we did was, we had test tube 1 and test tube 2, what we did was we weighed the test tube by itself then the test tube with the hydrate. Then we subtracted the test tube by itself that gave us the known. After we used the Bunsen burner to remove all of the H₂O we subtracted the test tube from the amount we got from that to get the experiment. Then we plugged it all into the formula and got the percent. The results are very precise (refers to numbers on board), but not that accurate.

T1’: We actually got the zinc (sulfate), I don’t know if it is right, but we got the zinc.

T2’: Did he (the instructor) say y’all’s is right?

P’: No, not yet.

T2’: I guess we will find out.

P’: Do y’all have your numbers here? (Takes paper from T2.) Yeah, I think you guys probably did get zinc.
T2’: Yeah, we had one that was like 45% and one that was…

T3’: Y’all got copper sulfate, who else got copper sulfate?

T1’: They did and they did (pointing around the room). The other groups that switched had copper, no one else has zinc.

T3’: But one of them could have got it wrong.

T1’: I think he (the instructor) will say at the end who was right. (Instructor comes by at this point.) Are we supposed to have the same hydrate, because we have different answers?

I: Yes, so somebody is wrong.

T1’: Does that mean we will fail our lab report?

This discussion was focused entirely on deducing the correct answer by eliminating hydrates identified by other groups. There was no discussion of the data, methodology or even the calculations. The instructor unfortunately only exacerbates the situation by declaring that “someone is wrong”. A better response would have guided the students to consider why the water percentage would be low or even to suggest that they examine the data generated by each group.

The following conversation took place after the argument evaluation. The excerpt picks up with the presenter from table 6B presenting a forced conjecture as to why his group’s hydrate was not blue like other groups that had a copper salt.

P’: There were three people that had this and they all had a blue one, be we don’t know if they dyed it or not, so we have to find out if we are right. (Instructor comes over and talks for sometime about how to do the calculations correctly.)

P: (The presenter from table 2B comes over to talk.) I think it was the way you burned it, did you only heated the bottom?

P’: No, we heated up the side to get the water out. So, it still had some water in it?

P: Yeah, that is what I think.

T1: What should we do?
We have to just explain that man.

Several trends are evident in this series of conversations. The problem with table 6B’s data was that it still had water present. The presenter from table 2B suggested this possibility at the end of the lab, but the idea did not resonate with the group. The focus was on right versus wrong rather than evaluating the evidence. The students’ reliance on authority is clear in this situation as well as the tendency to use inappropriate reasoning to justify a claim, both of which are recognized barriers to productive scientific argumentation (Kuhn & Reiser, 2005; Sampson & Clark, 2008a). For example, the lack of blue color was actually a clear indication that the group did not have a copper salt, rather than using this to adjust their claim they decide that maybe there is dye in some of the samples. Finally, after all of the input from the other group and the instructor, the presenter for table 6B was caught on tape the following week saying, “We had a little error last week, but I just wrote what we did and that it was wrong.” The student’s lab report was a exact copy of his table’s whiteboard with the conclusion that his answer was wrong because the other group was right with a claim of zinc sulfate. He had included none of the corrections that came out of the argument evaluation or the discussion with the instructor, but stuck to his original claim.

This example was used, even though it was early in the semester, because of the conversation that was caught on the tape at the end of lab the following week. In fact this situation occurred repeatedly over the course of the semester. This is a disturbing observation given that the ADI instructional model is designed to create a scientific community in both the collaborative groups and through the poster sessions; yet, the written arguments consistently disregarded or minimized the evidence presented by other groups. Field observations found that the students operated under an assumption that using another table’s data to support a claim or to support a claim change violated an unspoken code of ethics (i.e., cheating). The students’ struggled with accepting that they would not be told if they were “right”, rather that they had to make a strong argument supported by evidence from multiple sources, including peers.

This finding has ramifications for the teaching and learning of the Nature of Science, specifically the tenet that science is a social practice and that scientific knowledge is the product of community consensus (Abd-El-Khalick & Lederman, 2000). One of the implicit goals of the ADI instructional model is to foster an understanding that science is a social activity, not just in
the sense of society determining the science that gets done, but also that the rules of scientific procedure and the legitimacy of the ‘product’ are determined by a community of practitioners (Hodson, 2008). In addition, there are those that suggest that without claim change then there is no argument (Berland & McNeill, 2010). This seems a narrow view of argumentation, however not making a justified claim change or offering a rebuttal is indicative of a failure to understand an overarching premise of argumentation and thereby its role in science.

Two weeks later during the argumentation session for the solutions laboratory, there was a similar situation where the tables that had the same unknown had different molar concentrations for the two dyes in their solution. Rather than just accepting that someone was right or wrong, this time the students tenaciously compared their data and their calculations until they found an error. The group of students with the error corrected their concentration values on their whiteboard and used the corrected values in their laboratory reports as well. This was one of very few instances of claim change following the argument evaluation session. This occurred on the third ADI investigation suggesting that the role of argumentation was being accepted by the students and at least these two tables were participating in classroom culture that was comparable to the culture of science.

Over time the arguments improved in both use of reasoning and willingness to change claims. A particularly illustrative example occurred during the argument evaluation for the mole ratio investigation. This is an investigation were everyone could come to the same conclusion and most groups did agree that the mole ratio is 1:1. In this example, the presenter for table 5A adamantly moves the discussion from claim to reason.

T1:  *(looks at board)* Yeah, we got the same thing.

P:  Yeah, ok, but let’s see if we had the same reasons. We said 1:1. *(referring to graph)* We thought that since it was increasing by so much for the first three and then it started leveling off after so that adding moles after you got the the 1:1 ratio, didn’t make much more of the CO$_2$. For the 1:1, we got 380 mL, if we did it twice we would get 760 mL. Now, this one uses twice as much and you only get 420.

T1:  That makes sense.

T2:  A lot of people are saying, the 1:2.

P:  I think once the people that got 1:1 start explaining, they will switch over because it just makes sense.
The statistical evidence for growth, which these examples illustrate, supports the conclusion that the ADI instructional model can impact the argumentation that takes place between undergraduate students over the course of a semester. These results suggest, at the minimum, that the notion of claim changes in light of evidence obtained by others was new to students as well as the concept of moving beyond right and wrong to a reasoned explanation.

**Question 4 – ASAC as Predictor of Written Argument**

*To what extent does the quality of the scientific argumentation that takes place between undergraduate students predict written arguments for individual undergraduate students?*

Given the similar patterns for the written argument scores and the two ASAC scores, the significant correlations were predictable. The significant correlation for the two ASAC scores suggests that similar levels of argumentation can be obtained in both collaborative groups that develop a level of comfort with each other and in small random groups that varied in their composition for each event. In planning this study the hypothesis was that the ASAC might reveal distinguishing features of the two types of argumentation events, however, in reviewing the scores for the subcomponents of the ASAC, this was not apparent in this study. Again, the small N value limited this type of assessment.

The level of argumentative discourse as indicated by the ASAC score was predictive of the written argument score, explaining 12% of the observed variations in the writing scores. This finding is consistent with the general thesis that participation in oral argumentation can transfer to the individual’s written argument. Other researchers attempting to connect oral and written discourse have found that written arguments are typically not as robust or complex as the oral argumentation (Berland & McNeill, 2010; Kantor & Rubin, 1981). Berland & McNeill (2010) suggest that the presence of an audience during an episode of argumentation forces students to develop rich, convincing arguments. Without the audience, the written argument may become more of a demonstration of content understanding to the teacher, prompting the authors to suggest that different instructional support may be required to develop better skills in these two contexts. This study did not specifically address a “lag” in written argument relative to oral argumentation; rather the intention was to look for growth in both contexts over the course of a semester. Significant growth in both oral argumentation and written argument were observed with the additional finding of a significant correlation between the two contexts. Given the
variation that occurs within different types of argumentation events, this finding suggests that the ASAC observation protocol may be useful not only in scoring oral arguments but in predicting the likely transference to written arguments as well.
CHAPTER 5  
LIMITATIONS, IMPLICATIONS AND CONCLUSIONS  

Limitations  

In recognition of the limitations of a small $N$ value, the results of this study should be interpreted cautiously. That being said, sample size would be a greater concern in a direct comparison study. In a longitudinal study, in contrast, including a smaller number of participants is appropriate because multiple measures are taken over time. So while the $N$ value for each data source is approximately 30, the analyses required the researcher to score 150 laboratory reports, 90 argumentation events and 90 performance-based assessments. The similarity in the trend for the ASAC and written arguments also strengthens the claims for significance, as does the similarity in the growth of the various components of the performance-based assessment.

The participants in this study were a diverse group that is representative of a general undergraduate population in terms of gender, race and academic ability; yet this sample also presented challenges for research in terms of student accountability. Inconsistencies in attendance and assignment submission ultimately reduced the data available and limited the scope of the analyses. Given larger $N$ values, a microanalysis of the data might have provided more insight regarding how different subgroups (gender, race or GPA) benefit from this type of instruction. Analysis of the components of the ASAC and the written argument would have been informative, but were precluded by the sample size. Future research might be able to address these issues.

An additional limitation of this study was a potential teacher effect. The instructor who taught the sections of the lab course was not as experienced as one might hope for an ideal implementation of this instructional method although he did move each of the students through each stage of ADI with a high degree of fidelity. A recognized barrier to inquiry-based instruction is the need for experienced instructors (Driver et al., 2000). This instructor was not a strong classroom manager and he was not well organized in getting the investigations started. On the other hand, there was still growth in all aspects of the study even given the contextual limitations. Finally, the extent to which the presence of researchers, video cameras and an atypical atmosphere in general influenced events cannot be calculated, but there almost certainly was an impact.
While this study did not have a comparison group, previous research (Walker, Sampson, Grooms, Zimmerman et al., in press) did directly compare the ADI instructional method with more traditional laboratory instruction. The results of that study showed a significant difference between the comparison group and ADI group in ability to use evidence and reasoning to support a claim. While a comparison group would have been valuable for assessing directly the degree to which the ADI instructional model contributed to the observed growth, it was not critical in this study because I was interested in looking at growth over time. In addition, undergraduate laboratories are typically not designed with argumentation opportunities like those in ADI so it would have been difficult to identify legitimate comparison scenarios. Students learning the expectations of the model and simply adapting to the requirements could also help explain the observed growth. However, if this was all that was required then I think there would have been less of a spread in the data as the semester progressed. I think that despite becoming comfortable with the instruction and gaining in content knowledge, students still struggled to justify a claim with sound evidence and reasoning. It will take more than five events for some students to learn how to craft a high quality argument and participate productively in an episode of scientific argumentation. With these limitations in mind, I will now outline the conclusions that I have been able to draw from the data that was collected during this study.

**Implications for Instruction**

Although there has been much research on argumentation in science in recent years, little has been situated at the post-secondary level. There is primary school research (Reznitskaya et al., 2001), a large body of middle school research (Berland & McNeill, 2010; Osborne et al., 2004) and some secondary school research (Jimenez-Aleixandre, Rodriguez, & Duschl, 2000; Sampson & Clark, 2009). In addition, the research that is available is often limited to pre/post assessment designs (Zohar & Nemet, 2002) that measure absolute change, but do not examine the process of change. This is to some extent due to the limited time frame involved; with most interventions lasting a few days or weeks. While these studies have found significant differences between control and experimental groups on the outcome of interest as well as for improvement over time as measured by pre/post assessments (Nussbaum, Sinatra, & Poliquin, 2008; Osborne et al., 2004; Walker, Sampson, Grooms, Zimmerman et al., in press), there is less information on students’ growth in argumentation.
The ADI instructional model uses a sequence of activities – investigation design, oral argumentation, written argument, and peer review – in order to give a more central place to argumentation. These activities provide opportunities for explicit instruction in argumentation, while at the same time creating classroom contexts that promote and support student engagement in argumentation. This instructional environment made it possible to evaluate growth in argumentation over time and in multiple contexts.

**Method-focused vs. Claim-focused Argumentation**

The significance of each of the measurements used in this study has been discussed, but the similarity in trends across the oral and written discourse and what the findings might say about the classroom context created by the ADI instructional model is an important factor that warrants discussion. The oral argumentation and written argument scores in the first laboratory investigation were much higher than expected, as noted earlier, this can be attributed to the introduction of an unfamiliar tool, the spill can, for measuring water displacement. This unfamiliar tool resulted in a different type of argumentation that was more focused on the method or the tools used to collect data than the claim and the evidence for it. The use of the spill can and the erroneous results that were obtained by the various groups resulted in great deal of argumentation that was focused almost entirely on measurement and methodology. This was in sharp contrast to the investigation that followed, the identification of unknown hydrate, where the students’ argumentation was much more focused on the content of the claims and how well the claims fit with the available evidence, even though many of the groups obtained erroneous results.

The argumentation in the next three laboratory investigations also focused more on the content of the claims and how well they were supported by evidence than it did on measurement or methodology. As the students gained experience with oral argumentation, then their argumentation showed the linear growth that was described earlier. Finally, in the last investigation, there appeared to be little disagreement between the various groups because they all reached similar conclusions. As a result, there was little discussion devoted to the nature of the claim, how well the evidence fits with a claim or methodology because almost every group had the same answer. The nature of the task and students’ familiarity with lab technique, therefore, appears to impact the type and level of argumentation observed and can result in argumentation that is either methodology-based or claim-based.
Given this finding, science educators will need to be mindful of the nature of the task that is used to promote and support student engagement in argumentation. If the research or instructional goal is to have students generate data that they can use to develop a claim, support it, and evaluate the claims of others, then the tools presented for data collection and the methods suggested should be familiar ones. On the other hand, providing students with unknown materials for them to use during an investigation may result in argumentation that is more focused on methodology. Although neither approach is inherently better than the other in terms of promoting and supporting student engagement in argumentation, science educators must be aware of how the context of the task can influence the focus of an episode of argumentation.

**Integrated vs. De-contextualized Instruction**

The results of this study also contribute, albeit indirectly, to the on going debate about how instructors should help students learn how to participate in scientific argumentation and craft written arguments. Authors such as Simon, Erduran and Osborne (2006), for example, suggest that people need to learn what counts as an argument and how to participate in argumentation as a precursor to learning content or doing science. This position takes a more process- or mechanistic-orientated approach to learning and thus requires students to learn the rules of argument and argumentation before they are allowed to learn the concepts, theories, and models of science. Sandoval and Millwood (2007), on the other hand, argue that educators should embed argumentation within inquiry-based instruction so students can learn how to craft an argument and participate in argumentation over time while they learn content. From this position, argument and argumentation are viewed as learning tools and there is no separation between these processes and the learning of content. Indeed, this enculturation of the tools and symbols of science is the goal of an instructional model framed from a social constructivist perspective. As students repeatedly use the tools and processes of science they develop criteria for “what counts” in science. In argumentation contexts, “students become active producers of justified knowledge claims and efficient critics of others’ claims” (Jimenez-Aleixandre, 2007, p. 98).

While there is much debate about the relative merits at the extremes of these two positions, I see value in each and think students need to engage in argumentation in order to develop a better understanding of the content (i.e., arguing to learn) but also students need to learn what counts as an argument and argumentation in science (i.e., learning to argue) as part of the process. The ADI instructional model is well aligned with this more integrated perspective.
These findings, therefore, demonstrate that science instructors can, at least in this type of context, accomplish both tasks (i.e., arguing to learn by learning to argue) at the same time.

**Peer-review as Audience**

Peer-review is a process used to establish rigor in scientific research and thus was an important component of the ADI instructional model. McNeill and Krajcik (2008) suggest that the absence of an audience in the written argument contributed to their conclusion that student written arguments were more content based relative to oral arguments. This concept is intriguing in light of the role that peer-review could play in providing an audience. In the ADI instructional model, students write laboratory reports based on a rubric and then peer-review each other’s reports using guidelines taken from this rubric. McNeill and Krajcik (2007) suggest that different instructional support may be required to develop better written argument skills. This study found that oral and written arguments developed along similar trends and were in fact significantly correlated. The extent to which these results can be explained by the peer-review process is worthy of further consideration. From this research I would suggest that the peer-review and shared criteria for “what counts” in a written argument contributed to the observed growth in written arguments and to the significant correlation with oral argument. Research to specifically investigate this claim is in progress.

**Impact on Presenter**

The research data and subsequent analyses were focused on examining the impact of the ADI instructional method on oral and written argumentation. Other areas that might have been studied would be the process of investigation design and the discourse that takes place during the peer review. It might be useful to develop a means of scoring the whiteboards developed for the poster sessions as they provide a snapshot of the group’s evidence and rationale that could be compared to their written arguments. In watching the videos, I was struck by how much the presenter’s explanations and reasoning improved with each presentation of the material. They would learn from the conversations and clarify their own understanding. The impact for this individual would be interesting to evaluate as well.
Conclusions

Argument and debate are essential processes in science and are linked to understanding of science, yet there is little emphasis on the role of argumentation in science education (Osborne, 2010). It was the intent of this research to add to the growing body of information that demonstrates the importance and value of instructional models that engage students in argument and critique, not only for understanding of science content but for developing science literacy as well (Sampson & Clark, 2006, 2008b, 2009; Sampson et al., 2009). The added significance of this research was that ADI was implemented and studied in an undergraduate science laboratory program, where the use of innovative instructional methods is rare but where the impact of current methods of instruction have an overarching impact on the entire education system (NRC, 1999; Sunal et al., 2004). The significant growth in both oral argumentation among groups and in written arguments for individuals that was documented in this study indicates that ADI can help undergraduates learn how to participate in these complex scientific practices.

The social constructivist foundation for learning that ADI was situated upon has been referenced repeatedly in this document. The extent to which this research supports the notion that learning science should be viewed as process of social participation (Cobb, 1994; Driver et al., 1994) warrants explicit discussion. The ADI instructional model provides students with opportunities to participate in the processes of science. Students generate products or answers in the form of proposals, claims, solutions, and experimental designs. Students choose between competing explanations and back their claims with evidence. They use shared criteria to evaluate evidence and report on their conclusions. Students talk science, they write science and they participate in a context that models the culture of science to the fullest extent possible in an undergraduate lab course. The results of this study demonstrate that participation in this culture results in students’ incorporation of the aspects of quality of argumentation into their discussion, writing and individual assessment. This occurred with little explicit instruction in how to argue, but developed through social participation.

The results of this study, however, do not mean that science educators interested in helping students learn to participate in scientific argumentation need to adopt and use the ADI instructional model exactly as described in this research or in other studies. There are numerous ways that science educators can adapt existing instruction in order to promote argumentation. For example, the physics instructors at the community college have incorporated the whiteboard
presentations into the laboratory curriculum. This relatively simple change has enhanced student understanding of the variability of results and improved the data analysis in their laboratory reports (Carr, 2011).

Kuhn (1991) suggests that for the overwhelming majority of students, the use of valid argument does not come naturally and is acquired only through practice. The ADI instructional model with its repeated opportunities to develop arguments and to engage in argumentation appears to be a means for post-secondary faculty to place value and emphasis on these processes. It is through this experience that students can come to understand science as a way of knowing and thereby advance the cause of science literacy.
APPENDIX A
ADI INVESTIGATIONS
Lab Investigation #1: What is the best practice for laboratory measurement? Using Laboratory Equipment and Glassware

1. The chemistry laboratory is equipped with top loading balances and analytical balances. On your lab table you will find an object. We would like you to weigh the object on both types of balances. Do this individually using a different set of balances so that you can compare values with your partner.

<table>
<thead>
<tr>
<th>Student</th>
<th>Top Loader #</th>
<th>Analytical #</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Record all decimal places and units.*

a. How do the values for Top loader vs. Analytical compare?

b. How do values compare between students? Identify some reasons for any differences.

2. Measure 8 mL of water in the 10 mL graduated cylinder and in the 25 mL graduated cylinder.

a. To how many decimal places can volume be measured with the 10 mL graduated cylinder?

   Circle one: 0.1 0.01 0.001

b. Write the volume using the correct number of decimal places and units. ____________

c. To how many decimal places can volume be measured with the 25 mL graduated cylinder?

   Circle one: 0.1 0.01 0.001

d. Write the volume using the correct number of decimal places and units. ____________

3. Measure 10 mL of water in the small beaker provided, then transfer the water to the 10 mL graduated cylinder (smallest one). Read the graduated cylinder to two decimal places.
a. What volume do you have in the graduated cylinder? ________________
   *Record all decimal places and units.

b. Why would the volume be different when measured with the 10 mL graduated cylinder versus the beaker?

c. What can you infer about the volume accuracy using a beaker?

4. Use the ruler and calipers provided to measure the length of the object on your lab table. Do this individually.

<table>
<thead>
<tr>
<th>Object Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Ruler</td>
</tr>
</tbody>
</table>

*Record all decimal places and units

a. Technically you can measure to the 0.01 with both the ruler and the caliper. Why would you choose to use the calipers?

b. How do values compare between students? Identify some reasons for any differences.

5. What is the best practice for laboratory measurement?
Lab Investigation #2: Are these materials made of the same substance? An introduction to physical properties.

Introduction: A physical property can be observed or measured without changing the identity of the material. This is in contrast to a chemical property which does change the identity of the material. For example, the difference between melting iron which is a physical change and the reaction of iron with oxygen to form rust which is a chemical change. Typical physical properties include shape, color, odor, density, hardness, melting point and boiling point. Density is a property that is often used in chemistry because it is a useful conversion factor between mass and volume. Density is also, easily measured in the lab and can be used to distinguish different substances.

The Problem: Matter, the “stuff” of which the universe is composed, has two characteristics: it has mass and it occupies space. The relationship between these two characteristics is called density. Since density is a physical property of matter, density can be used to determine the identity of a substance. In this investigation you will attempt to identify several unknown samples of matter based on their density.

The guiding question of this investigation is: Are the samples provided made of the same material?

Materials available for use: You may use the following materials during your investigation.

- Water
- Spill Can
- Tongs
- Graduated cylinders
- Beaker
- Top Loader Electronic Balances
- Vernier Calipers

Safety Precautions:
- Carefully add the samples into the glass cylinders to avoid cracking the glass.
- Wear goggles at all times.

Getting Started: You will need to determine a way to gather the data so that you produce evidence that can be used to justify your answer to the guiding question. It is good scientific practice to determine values in more than one way, so you should develop two methods for determining the density of your samples. You will then coordinate these components into an argument that you can use to convince your classmates that your ideas are valid and acceptable.

Interactive Poster Session: Once your group has completed your work, prepare a whiteboard that you can use to share and justify your ideas. See the handout provided for details on this process.

Report: Once you have completed your research, you will need to prepare an investigation report that consists of three sections. Each section should provide an answer for the following questions:

Section 1: What were you trying to explain (or figure out) and why?
Section 2: How did you go about your work and why did you conduct your investigation in this way?
Section 3: What is your argument?

Your report should answer these questions in 2 pages or less. This report must be typed and any diagrams, figures, or tables should be embedded into the document. Generally, you need one page for the first two sections and the second page for third section. The third section is where you not only present your data, but use the values you obtain as evidence in your reasoning. Be sure to write in a persuasive style; you are trying to convince others that your explanation is acceptable or valid! Statements like, “see data table for values” are not be acceptable!
Lab Investigation #3: What is the identity of this hydrate? Identification of Hydrates Based on Percent Water in an Unknown

Introduction: When an ionic compound (such as salt) is crystallized from an aqueous solution (such as salt water), the resulting solid crystals may appear to be perfectly dry. When the crystals are heated, however, the mass of the solid may decrease as water is released from the crystal structure. The form or appearance of the crystals may change and, in some cases, the color of the crystals may also change. Compounds that contain water molecules as part of their crystal structure are called hydrates. Are hydrates pure substances or are they simply “wet salts,” that is mixtures containing variable amounts of water?

A hydrate is a pure substance because it contains water molecules embedded in its crystal structure that do not vary. Heating a hydrate “drives off” the water molecules, and the solid that is left behind is called anhydrous (which means “without water”). The chemical formula of hydrate specifies the relative number of each kind of atom in a molecule, as well as the number of water molecules bound to each molecule. Calcium chloride dihydrate (road salt) is an example of a hydrate. The chemical formula for calcium chloride dihydrate is CaCl$_2$ • 2H$_2$O. The “dot” in the chemical formula indicates that two water molecules (H$_2$O) are attached or bound to the calcium chloride (CaCl$_2$) ions by weak chemical bonds. The water molecules in calcium chloride dehydrate can be removed by heating the hydrate (see below).

\[
\text{CaCl}_2 \cdot 2\text{H}_2\text{O} \xrightarrow{\text{heat}} \text{CaCl}_2 + 2\text{H}_2\text{O}
\]

The number of water molecules in a typical hydrate is characteristic of the particular salt and is usually a small whole number from 1 to 10. The chemical formula of a hydrate can be determined by analyzing the percent water in the hydrate, the ratio of the mass of water lost upon heating divided by the mass of the original hydrate. The formulas of some common hydrates and their anhydrous salts are summarized below.

<table>
<thead>
<tr>
<th>Common Name</th>
<th>Chemical name</th>
<th>Hydrate</th>
<th>Anhydrous Salt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Washing Soda</td>
<td>Sodium Carbonate</td>
<td>Na$_2$CO$_3$ • H$_2$O Na</td>
<td>2CO$_3$</td>
</tr>
<tr>
<td>Gypsum Calcium</td>
<td>Sulfate</td>
<td>CaSO$_4$ • 2H$_2$O CaSO</td>
<td>4</td>
</tr>
<tr>
<td>Epsom Salt</td>
<td>Magnesium sulfate</td>
<td>MgSO$_4$ • 7H$_2$O MgSC</td>
<td>4</td>
</tr>
</tbody>
</table>

The Problem: You will be given an unknown hydrate to identify this hydrate from a list of possible unknowns by determining the percent water in the hydrate.

The guiding question for this investigation is: What is the identity of your unknown hydrate?

Safety Precautions:
- Never point the heating test tube at a person.
- Only heat to temperature needed to remove water, excess heat will result in oxidation of the salt.
- Wear goggles at all times.

Materials available for use:
- Test tubes
- Test Tube clamp
- Beaker
- Top Loader Electronic Balances
- Bunsen Burner
- 0.5 to 1.0 gram of unknown hydrate
**Getting Started:** The unknown hydrate may be any of the compounds in the table on the design page. To determine the amount of water found in a hydrate experimentally, you must remove the water from the hydrate. To do this, you will need to heat the hydrate with a flame in order to evaporate the water (see Figure 1).

![Figure 1: How to evaporate water from a hydrate](image)

**Technique Considerations:** Take a minute to discuss why both of these are important with your group.

- Be sure to weigh your empty test tube, so you don’t lose material transferring back and forth. Stand the test tube in the beaker after zeroing the balance with the beaker on the pan.
- Spread the salt up the side of the test tube to increase surface area.

**Interactive Poster Session:** Once your group has completed your work, prepare a whiteboard that you can use to share and justify your ideas. See the handout provided for details on this process.

**Report:** Once you have completed your research, you will need to prepare an investigation report that consists of three sections. Each section should provide an answer for the following questions:

Section 1: What were you trying to explain (or figure out) and why?
Section 2: How did you go about your work and why did you conduct your investigation in this way?
Section 3: What is your argument?

Your report should answer these questions in 2 pages or less. This report must be typed and any diagrams, figures, or tables should be embedded into the document. Generally, you need one page for the first two sections and the second page for third section. The third section is where you not only present your data, but use the values you obtain as evidence in your reasoning. Be sure to write in a persuasive style; you are trying to convince others that your explanation is acceptable or valid! Statements like, “see data table for values” are not be acceptable!
Lab Investigation #4a: Spectroscopy and Chromatography – Two Key Techniques for Identifying Chemical Compounds

**Introduction:** Two powerful techniques used by scientists are spectroscopy and chromatography. There are many forms of each technique. In general, spectroscopy uses the interaction of chemical molecules with light to characterize a substance. We will see that we can use the absorption of light to determine concentration. Chromatography is primarily a method of separating mixtures of compounds in order to identify the individual components. We will use a simple form of chromatography to determine the components of an unknown mixture of food dyes. Week I of this lab activity will provide you with the basic knowledge and skills to use these two techniques. In Week II, you will put your new skills to the test and identify not only the components of the mixture but the concentration of each component.

**Week 1: Spectroscopy and Chromatography – How to guide.**

**Spectroscopy Background:** Begin by viewing the online video “Spectronic 20.” In this video background information on the theory and use of a spectrophotometer is provided including how to calibrate a spectrophotometer, generate a Beer’s Law plot and use it to determine the concentration of an unknown.

There are two principal ways in which the spectrophotometer displays data. The %T is a linear scale in which 100% T means that the sample is absorbing none of the light passing through it to 0% T in which all of the light is being absorbed (0% is making it through the sample). The Spectronic 20D most accurately measures light in the percent transmission mode. While %T is easily measured, the absorbance is a more useful value. The conversion between %T and absorbance is a simple one.

$$A = 2.000 – \log(\%T \text{ sample})$$  \hspace{1cm} \text{Equation 1}

We relate absorbance of the sample to concentration by using the Beer-Lambert law (also known as Beer’s law). The Beer-Lambert law states that the absorbance of a sample depends on the product of the concentration of the sample (c, the Molarity of the sample) by the path-length (b, the distance the light travels through the cuvette in cm) by the molar absorptivity (ε, which is a correction factor taking into account the sensitivity of the sample to the wavelength chosen).

$$A = \epsilon bc$$  \hspace{1cm} \text{Equation 2}

If we use the same cuvette for all the studies and study the same system at the same wavelength, b and ε are constant values. Therefore, absorbance is directly related to concentration and a linear relationship is observed. See Figure 1.

Once a Beer’s Law plot (calibration curve) is generated for a substance using known concentrations, unknown sample concentrations can be determined by measuring the sample’s absorbance and comparing to the calibration curve.

71
For example, if the absorbance was measured as 0.65, then the concentration for this particular substance, based upon its calibration curve, is $22 \times 10^{-5}$ M or $2.2 \times 10^{-4}$ M. It is not always this straightforward to find the concentration by reading off of the graph. More often we use the equation for a straight line, $y = mx+b$, where $m$ is the slope and $b$ is the y-intercept.

In this case $m = 0.030$ and $b = 0$. Using these values the equation becomes: $y = 0.03x + 0$. Substituting the absorbance value: $0.65 = 0.030x$; $x = 21.6 \times 10^{-5}$ M or $2.2 \times 10^{-4}$ M.

**Determining an Analytical Wavelength from Visible Spectrum of Dyes**

1. Obtain two cuvettes for the UV-Vis Spectrophotometer. Fill one cuvette about ¾ full with deionized/distilled water to serve as a reference (blank). Fill the second cuvette with dye solution about ¾ full. Your team will be provided one of the following three FD&C dyes: Blue 1 (Brilliant Blue), Yellow 5 (Tartrazine), or Red 3 (Erythrosine).

2. Make sure to handle the cuvette by holding at the rim (mouth) of the cuvette and not in the path of the light. Finger grease and other materials will absorb light and throw off your results. Clean the outside of the cuvette with a KimWipe (or other lint-free) tissue.

3. Double-click on the Lambda 25 icon to launch program. Select the file **FOODDYES.MSC**.

4. Insert the water (blank) cuvette in the reference well (rear compartment) and click start. The program will tell you to place the sample (dye) cuvette in the sample well (front compartment). Make sure that the cell is inserted so that the longer path length (1 cm) is in line with the light beam. Check with instructor to verify correct placement. Click OK to begin scan.

5. Once the scan is complete, see Figure 2 below for instructions on how to label the spectrum with the name of dye and the peak with wavelength. Note: this peak, the maximum absorbance, is the analytical wavelength for the dye that will be used to set the Spectronic 20D for obtaining data for the Beer’s Law plot.

6. After labeling dye and peak wavelength, choose **print** from under **File** menu. Print 6 copies, one for each team member.
Preparing Food Dye Solutions for the Beer's Law Plot

(1) Label 5 beakers, stock and “A” through “D” and fill the stock beaker to about 30 mL with the stock dye solution assigned to your team. 
(2) Obtain two 10-mL graduated cylinders. Label one graduated cylinder dye and the other deionized/distilled water.
(3) Rinse the dye graduate cylinder with the stock solution then use the dye graduated cylinder to transfer 10.00 mL of the stock dye solution into beaker A.
(4) With the deionized/distilled water graduated cylinder, add 10.00 mL of deionized/distilled water to beaker A.
(5) Repeat this procedure three more times using solution A to prepare solution B and solution B to prepare solution C, and solution C to prepare solution D. Each time rinse the dye graduated cylinder with 1 mL of solution being diluted.

This is known as a serial dilution.
Preparing a Beer’s Law Plot

(1) Obtain two “matched” cuvettes for the Spectronic 20D. Fill one cuvette (sample) about ¾ full with the dye from beaker D, and the second cuvette with distilled/deionized water (reference/blank). This reference/blank is usually the solvent by itself or a solution of all the components other than the specific solute under study and will be used to set the spectrophotometer to 100%T.

(2) Set the Spectronic 20D to the appropriate analytical wavelength for the dye to be measured. Set the output to % Transmittance.

(3) With no cuvette in the sample holder and the sample lid closed, use the front left knob to set the readout to 0 %T.

(4) Insert the blank into the sample well making sure the marking on the cuvette is lined up with the notch on the sample well. Use the front right knob to adjust the readout to 100 %T. Make sure to only handle the cuvette at the rim (mouth) and not in the path of the light. Finger grease and other materials will absorb light and throw off your results. Clean the outside of the cuvettes with KimWipe tissue.

(5) Remove the blank and insert the sample cuvette with solution D, record the %T. Remove sample cuvette and discard dye into waste beaker.

NOTE: The spectrophotometer must be calibrated to 100% and 0% before every measurement because the spectrophotometer’s accuracy varies with time, current fluctuations and temperature changes. This is why you need to keep your reference cuvette in a safe place to re-calibrate the machine often. Repeat measurement procedure using solutions A, B & C.

(6) Convert %T to Absorbance using Equation 1. Use Graphical Analysis (or another spreadsheet program) to plot Absorbance versus concentration (Beer’s Law Plot).

(7) Perform least-square analysis (linear regression) to obtain calibration curve. The closer the correlation coefficient (R²) is to 1, the more linear the data. Look at your graph, if there is an obvious outlier, you may want to repeat that measurement.

Practice: Determining unknown concentration of dye

(1) Obtain a practice solution of the same FD&C dye that was used to prepare the calibration curve.

(2) Follow same calibration procedure to set 0%T and 100%T; then measure %T of unknown concentration.

(3) Convert %T of unknown to Absorbance; then use the calibration curve to determine the practice solution’s concentration. Verify result with instructor.

Chromatography Background

The technique of chromatography is commonly used to determine the composition of materials. There are many forms of chromatography but all forms employ a stationary phase (a solid) and mobile phase (a gas or a liquid solvent). Separation and identification of a material relies upon the relative attraction of the various components of a mixture for the stationary versus the mobile phase. Substances that are more strongly attracted to the stationary phase will not move significantly; while substances that are more strongly attracted to the mobile phase will move up the paper as they travel with the solvent. In this lab, paper chromatography will be used in which the stationary phase is the paper and the mobile phase will be a 0.10% NaCl solution.

To determine the identity of a substance, the Rf value of a substance is compared to known substances. Rf values range from 0.0 (attracted 100% to solid phase) to 1.0 (attracted 100% to mobile phase). In order to
obtain the $R_f$ value for a substance, two measurements are needed: the distance the substance travels from the origin and the distance the solvent (mobile phase) travel.

In Figure 4, the $R_f$ value of dye 1 matches the $R_f$ value of one of the components of the mixture. To verify identity, the $R_f$ values should match. The $R_f$ value is calculated as follows:

$$R_f = \frac{\text{distance component travels}}{\text{distance solvent front travels}} = \frac{2.0 \text{ cm}}{6.0 \text{ cm}} = 0.33$$

**General Chromatography Procedure:**

1. Draw origin line on chromatography paper approximately 1 cm from bottom using pencil.
2. Place a sufficient amount of the chromatographic solvent (0.10% NaCl solution) in the bottom of a 250 mL beaker and place cover on the beaker to allow saturation of air. The chromatographic solvent should be below origin line.
3. 0.5% concentrated solutions of the known (B1, R40 etc…) dyes plus the unknown for Week 2 should be run on the same chromatogram to obtain consistent data. Below origin line, note identity of substances to be run making sure that there is sufficient distance between each.
4. When spotting substances on origin line use a clean applicator (toothpick) for each substance and spot several times waiting between applications for sample to dry. Note: The goal is a small concentrated spot of substance so repeated small applications is preferred.
5. Roll chromatography paper into a cylinder and staple. The ends should touch but not overlap.
6. Place chromatography paper in center of beaker careful that it is not touching the sides and cover. Allow chromatogram to develop.
7. Remove chromatogram before solvent reaches top of paper (~1.5 cm from top) since solvent will continue to rise and may run off top of paper. If this occurs the data is not valid. (Why? Consider how identity is determined using $R_f$ values.)
8. Handle chromatogram carefully when wet as it will tear easily. Once solvent has stopped rising, remove staples and mark solvent line but allow chromatogram to dry before taking measurements for $R_f$ values.
Lab Investigation #4b: How could you make more of this dye?

The Problem: Your research team is given an unknown dye solution and asked to determine its composition; that is, to identify the components that are present and the molar concentration of each.

The guiding question of this investigation is: How could you make more of this dye?

Materials available for use:
- UV-Vis Spectrophotometer
- Spectronic 20D spectrophotometers
- 10 mL Graduated Cylinders
- 50 mL Beakers
- Chromatography equipment
- Concentrated 0.5% dye solutions R3, Y5, B1
- Beer’s law plots for all three dyes
- Visible spectra for all three dyes

Safety Precautions:
- Wear goggles at all times.
- Concentrated dyes may stain skin.

Getting Started: Obtain an unknown dye solution from the instructor. You will be given two solutions, one for spectroscopic analysis and a small amount of a very concentrated solution suitable for chromatography.

Interactive Poster Session: Once your group has completed your work, prepare a whiteboard that you can use to share and justify your ideas. See the handout provided for details on this process.

Report: Once you have completed your research, you will need to prepare an investigation report that consists of three sections. Each section should provide an answer for the following questions:

Section 1: What were you trying to explain (or figure out) and why?
Section 2: How did you go about your work and why did you conduct your investigation in this way?
Section 3: What is your argument?

Your report should answer these questions in 2 pages or less. This report must be typed and any diagrams, figures, or tables should be embedded into the document. Generally, you need one page for the first two sections and the second page for third section. The third section is where you not only present your data, but use the values you obtain as evidence in your reasoning. Be sure to write in a persuasive style; you are trying to convince others that your explanation is acceptable or valid! Statements like, “see data table for values” are not be acceptable!
Lab Investigation #5: What is the optimum mole ratio for a reaction? An investigation of Mole Ratios and Limiting Reactants

Introduction: You have already learned how to balance chemical equations in terms of molecules, for example:

Unbalanced: \( \text{C}_3\text{H}_8 (g) + \text{O}_2 (g) \rightarrow \text{CO}_2 (g) + \text{H}_2\text{O} (l) \)

Balanced: \( 3 \text{C}_3\text{H}_8 (g) + 5 \text{O}_2 (g) \rightarrow 3 \text{CO}_2 (g) + 4 \text{H}_2\text{O} (l) \)

This equation can be interpreted as 1 molecule of \( \text{C}_3\text{H}_8 \) reacts with 5 molecules of \( \text{O}_2 \) to produce 3 molecules of \( \text{CO}_2 \) and 4 molecules of \( \text{H}_2\text{O} \). This information can also be interpreted in terms of moles (of molecules):

1 mole of \( \text{C}_3\text{H}_8 \) reacts with 5 moles of \( \text{O}_2 \) to produce 3 moles of \( \text{CO}_2 \) and 4 moles of \( \text{H}_2\text{O} \)

Why is this useful? With a balanced equation we can predict the moles of products that a given amount of reactants (in moles) will produce. When moles are used, we are then able to count the number of molecules produced by weighing (in grams). Predicting the amount of product formed or determining the amount of reactants needed for a reaction to occur is called stoichiometry.

Most stoichiometry calculations are performed using exact mole ratios of reactants and products. In real life, however, many commercial processes prepare compounds using an excess amount of one reactant (and thus a limiting amount of the other). For example, if you mix 2.5 moles of \( \text{O}_2 \) with 1 mole of \( \text{C}_3\text{H}_8 \), 3 moles of \( \text{CO}_2 \) will not be produced because there is not enough \( \text{O}_2 \) added to ‘use up’ all of the \( \text{C}_3\text{H}_8 \). Once the \( \text{O}_2 \) is consumed, no more products can be formed, even though some \( \text{C}_3\text{H}_8 \) remains. In this situation, because the amount of \( \text{O}_2 \) limits the amount of product that can be formed, it is called the \textit{limiting reactant} or \textit{limiting reagent}. Therefore, if two reactants are not mixed in the correct mole ratio, the reaction will not go to completion and you will have less product produced and one or more left over reagents.

The Problem: When bicarbonate is mixed with acid, it breaks down into \( \text{CO}_2 \) and \( \text{H}_2\text{O} \). Your task is to design and carry out an investigation to determine the optimum mole ratio for the formation of \( \text{CO}_2 \) by mixing various amounts of sodium bicarbonate and acetic acid. By comparing the amount of carbon dioxide generated when varying amounts of sodium bicarbonate react with a given amount of acetic acid, you should be able to determine the optimum mole ratio of sodium bicarbonate and acetic acid and be able to identify the limiting reactant in the other reactions.

The guiding question for this investigation is: \textit{What is the optimum mole ratio for the formation of} \( \text{CO}_2 \) \textit{from the reaction of sodium bicarbonate and acetic acid?}

Materials available for use:

- 1.00 M Acetic Acid (HC\(_2\)H\(_3\)O\(_2\))
- Sodium Bicarbonate (NaHCO\(_3\))
- Graduated cylinders (1000 mL & 25 mL)
- Plastic tray
- Electronic balance
- Beaker (400 mL)
- Side-arm flask w/ tubing
- Ring Stands/Rings
- Eye droppers

Safety Precautions:
- Handle acetic acid with care.
- Wear goggles at all times, as pressure is built up in this reaction.
**Getting Started:** You will need to collect gas by water displacement in order to measure the amount of CO$_2$ produced after the mixing acetic acid and sodium bicarbonate in different molar ratios. Once you have determined the amount of sodium bicarbonate you will need to use in each reaction, conduct your experiments. Be sure to keep in mind the goals of the investigation:

1. Determine the optimum mole ratio of sodium bicarbonate and acetic acid.
2. Write the balanced equation based on your data.
3. Identify the limiting reactant in each of the reactions.

**NOTE:** It may be helpful to prepare a graph of mL of CO$_2$ vs. moles of NaHCO$_3$.

**Interactive Poster Session:** Once your group has completed your work, prepare a whiteboard that you can use to share and justify your ideas. See the handout provided for details on this process.

**Report:** Once you have completed your research, you will need to prepare an investigation report that consists of three sections. Each section should provide an answer for the following questions:

Section 1: What were you trying to explain (or figure out) and why?
Section 2: How did you go about your work and why did you conduct your investigation in this way?
Section 3: What is your argument? Include a graph of mL of CO$_2$ vs. moles of NaHCO$_3$.

Your report should answer these questions in 2 pages or less. This report must be typed and any diagrams, figures, or tables should be embedded into the document. Generally, you need one page for the first two sections and the second page for third section. The third section is where you not only present your data, but use the values you obtain as evidence in your reasoning. Be sure to write in a persuasive style; you are trying to convince others that your explanation is acceptable or valid! Statements like, “see data table for values” are not be acceptable!
Lab Investigation #6a: Exploration of Chemical Reactions

Introduction: Chemical reactions can be recognized by color change, the formation of a solid, formation of bubbles, or a change in temperature. Chemists describe these reactions using chemical formulas. You have learned how to write and balance chemical reactions. But if we mix two or more reagents together, how can we determine what products are formed? In this investigation, you will determine the identity of the products that are formed as a result of a chemical reaction.

Exploration of Chemical Reactions: Carry out each of the reactions below. Note what physical states of matter are present in each reaction? Once your team has finished making the initial observations, use the solubility rules and splint tests for gases (see back of this page) to confirm the identity of the various products formed. Write balanced chemical reactions for the six reactions. Identify each reaction as Oxidation-Reduction, Precipitation or Acid/Base.

Reactions:
1. CuCl₂ (aq) + Al (s)
2. HCl (aq) + Zn (s)
3. CaCl₂ (aq) + AgNO₃ (aq)
4. HCl (aq) + NaHCO₃ (s)
5. H₂O₂ (aq) + catalyst
6. KNO₃ (aq) + Na₃PO₄ (aq)

Materials available for use: Test tubes & Rack
- Spatulas
- Rubber stoppers w/ vent
- Wood sticks
- Blue Litmus paper
- 3 M HCl
- 0.5 M CaCl₂
- 0.5 M CuCl₂
- 0.5 M KNO₃
- 0.5 M AgNO₃
- 0.5 M Na₃PO₄
- 3% H₂O₂
- NaHCO₃(s)
- Aluminum wire
- Zinc pieces
- Test tube with catalyst

Safety Precautions:
- Always wear goggles.
- Handle acid, HCl, with great care.
- AgNO₃ (silver nitrate) will stain your skin brown.
Lab Investigation #6b: Can you prepare 0.5 g of product?

Introduction: Last week you explored a variety of chemical reactions. It is valuable to know how chemicals react for many reasons; precipitation reactions are used to remove heavy metals from wastewater; and oxidation-reduction reactions are used to recover or plate metals. In both instances, we might want to recover or isolate the precipitate or the metal either for further use or disposal. This week’s lab provides the opportunity for you to use many of the skills learned in lab and lecture this semester. You will use your knowledge of chemical reactions and stoichiometry to prepare a specific amount of material.

The Problem: The second part of this investigation requires you to use your knowledge of chemical reactions from last week to design a procedure for producing ~0.5 gram of either Cu metal or AgCl (as assigned by your instructor).

Guiding Question: How do you know the reaction you selected formed the desired product and can you explain your yield in terms of the variables that we know affect chemical reactions?

Additional materials available for use:
- Side-arm filtration flask
- Büchner Funnel
- Filter paper
- 50 mL beaker
- 25 mL & 10 mL graduated cylinders
- Tweezers
- Tongs
- Micro spatula
- Watch glass
- Stirring rod w/ rubber tip

Interactive Poster Session: Once your group has completed your work, prepare a whiteboard that you can use to share and justify your ideas. See the handout provided for details on this process.

Report: Once you have completed your research, you will need to prepare an investigation report that consists of three sections. Each section should provide an answer for the following questions:

NOTE: Your lab report is to be written about the reaction for which you collected solid product.

Section 1: What were you trying to explain (or figure out) and why?
Section 2: How did you go about your work and why did you conduct your investigation in this way? Discuss your reaction design and show your calculations.
Section 3: What is your argument? You must justify your balanced equation with evidence you gather about the reaction. Your observations will be critical. Calculate a % yield for your product to use in your discussion.

Your report should answer these questions in 2 pages or less. This report must be typed and any diagrams, figures, or tables should be embedded into the document. Generally, you need one page for the first two sections and the second page for third section. The third section is where you not only present your data, but use the values you obtain as evidence in your reasoning. Be sure to write in a persuasive style; you are trying to convince others that your explanation is acceptable or valid! Statements like, “see data table for values” are not be acceptable!
APPENDIX B
MATERIALS FOR PERFORMANCE TASK
Limiting Reactants Answer Sheet

Each of the flasks contains the same amount of acetic acid (HC$_2$H$_3$O$_2$) and each balloon contains a different amount of sodium hydrogen carbonate (NaHCO$_3$). There is an acid/base indicator which will change from red (acid is present) to green (if no acid is present). The chemical reaction that occurs is given below.

\[
\text{HC}_2\text{H}_3\text{O}_2(\text{aq}) + \text{NaHCO}_3(\text{s}) \rightarrow \text{CO}_2(\text{g}) + \text{H}_2\text{O}(\text{l}) + \text{NaC}_2\text{H}_3\text{O}_2(\text{aq})
\]

<table>
<thead>
<tr>
<th>Balloon Color: ______________</th>
<th>Is the Limiting Reactant HC$_2$H$_3$O$_2$ or NaHCO$_3$ or Neither? (circle one)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>How do you know? Defend your explanation with appropriate evidence and a rationale.</td>
</tr>
<tr>
<td></td>
<td>The evidence is…</td>
</tr>
<tr>
<td></td>
<td>The rationale is…</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Balloon Color: ______________</th>
<th>Is the Limiting Reactant HC$_2$H$_3$O$_2$ or NaHCO$_3$ or Neither? (circle one)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>How do you know? Defend your explanation with appropriate evidence and a rationale.</td>
</tr>
<tr>
<td></td>
<td>The evidence is…</td>
</tr>
<tr>
<td></td>
<td>The rationale is…</td>
</tr>
</tbody>
</table>
# Limiting Reactant Performance Task Scoring Rubric

<table>
<thead>
<tr>
<th>Ratio of Reactants</th>
<th>Answer – 1 point each item</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2:1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limiting Reactant</td>
<td>NaHCO$_3$</td>
<td></td>
</tr>
<tr>
<td>Evidence</td>
<td>- The solution is red</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- The balloon barely inflated</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- There is no solid left in the flask</td>
<td></td>
</tr>
<tr>
<td>Rationale</td>
<td>- The color indicates that acid is still present</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- The solid is entirely used up so the solid is limiting</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- The small balloon indicates that more reaction is possible</td>
<td></td>
</tr>
<tr>
<td><strong>3:2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limiting Reactant</td>
<td>NaHCO$_3$</td>
<td></td>
</tr>
<tr>
<td>Evidence</td>
<td>- The solution is reddish/pink</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- The balloon is second smallest (or bigger than the red balloon)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- There is no solid left in the flask</td>
<td></td>
</tr>
<tr>
<td>Rationale</td>
<td>- The color indicates that acid is still present</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- The solid is entirely used up so the solid is limiting</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- The small balloon indicates that more reaction is possible</td>
<td></td>
</tr>
<tr>
<td><strong>1:1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limiting Reactant</td>
<td>Neither</td>
<td></td>
</tr>
<tr>
<td>Evidence</td>
<td>- The solution is orange</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- The balloon is maximum size</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- There is no solid left in the flask</td>
<td></td>
</tr>
<tr>
<td>Rationale</td>
<td>- The color indicates all the acid has been used</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- The solid is entirely used up so the solid is not in excess</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- The size of the balloon indicates that all the acid is used up</td>
<td></td>
</tr>
<tr>
<td><strong>2:3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limiting Reactant</td>
<td>HC$_2$H$_3$O$_2$</td>
<td></td>
</tr>
<tr>
<td>Evidence</td>
<td>- The solution is green</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- The balloon is the same size as the blue and orange balloons</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- There is solid left in the flask</td>
<td></td>
</tr>
<tr>
<td>Rationale</td>
<td>- The color indicates excess base (or that the acid is gone)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- The solid left in the flask indicates excess base</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- The size of the balloon indicates that all the acid is used up</td>
<td></td>
</tr>
<tr>
<td><strong>1:2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limiting Reactant</td>
<td>HC$_2$H$_3$O$_2$</td>
<td></td>
</tr>
<tr>
<td>Evidence</td>
<td>- The solution is green</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- The balloon is the same size as the blue and orange balloons</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- There is solid left in the flask</td>
<td></td>
</tr>
<tr>
<td>Rationale</td>
<td>- The color indicates excess base (or that the acid is gone)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- The solid left in the flask indicates excess base</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- The size of the balloon indicates that all the acid is used up</td>
<td></td>
</tr>
</tbody>
</table>

**TOTAL** /35
APPENDIX C

ASSESSMENT OF SCIENTIFIC ARGUMENTATION IN THE CLASSROOM (ASAC)
Assessment of Scientific Argumentation in the Classroom Observation Protocol

BACKGROUND INFORMATION

Teacher: __________________________ Announced? □ Yes □ No
District: __________________________ School: __________________________
Subject __________________________ Grade: __________________________
Observer: __________________________ Date: __________________________
Start time: __________________________ End time: __________________________

CONTEXUAL BACKGROUND AND ACTIVITIES
In the space provided below give a brief description of (a) the way the lesson was designed in an effort to promote and support argumentation and (b) the way the teacher managed the interactions that took place during the lesson. Be sure to note how often students are encouraged to engage in argumentation with each other in this classroom.

Describe the nature of the group (number of individuals, gender, race or ethnicity, native language, etc.) and the classroom setting (space, seating, etc.) in which the argumentation took place.
**CONTEXTUAL BACKGROUND AND ACTIVITIES**

In the space provided keep a running record of the events that occurred as the participants interacted with each other, the materials, and ideas.

<table>
<thead>
<tr>
<th>Time</th>
<th>Description of Event</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CONCEPTUAL AND COGNITIVE ASPECTS OF SCIENTIFIC ARGUMENTATION
HOW THE GROUP ATTEMPTS TO NEGOTIATE MEANING OR DEVELOP A BETTER UNDERSTANDING
(These items target how the group attempts to make sense of what is going on)

<table>
<thead>
<tr>
<th>1. The talk of the group was focused on solving a problem or advancing understanding.</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>

*Description:* The emphasis on advancing understanding indicates that there were some significant claims or explanations at the heart of discussion. Groups that score high on this item maintain the focus of their talk and efforts on understanding or solving the problem rather than the best way to finish their work quickly or with the least amount of effort. *Note:* Groups that stay on topic but never engage in an in-depth discussion about what is happening should be scored low on this item.

**Comments:**

<table>
<thead>
<tr>
<th>2. The participants sought out and discussed alternative claims or explanations.</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>

*Description:* Divergent thinking is an important part of scientific argumentation. A group that meets this criterion would talk about more than one claim, explanation, or solution. Individuals that valued alternative modes of thinking would respect and actively solicit new or alternative claims, explanations, or solutions from the other participants. *Note:* Groups that discuss multiple types of grounds or support for a claim, explanation, or solution but only one claim, explanation, or solution should be scored low on this item.

**Comments:**

<table>
<thead>
<tr>
<th>3. The participants modified their explanation or claim when they noticed an inconsistency or discovered anomalous data.</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>

*Description:* Inconsistencies between claims or explanation and the phenomenon under investigation are common in science. A group that modified their claim or explanation when they noticed inconsistencies or anomalies would not ignore “things that do not fit” or attempt to discount them once they are noticed by one of the participants. Groups that score high on this item try to modify their claim or explanation (not just their reasons) in order to account for an inconsistency or an anomaly rather than attempting to “explain them away”.

**Comments:**

<table>
<thead>
<tr>
<th>4. The participants were skeptical of ideas and information.</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>

*Description:* During scientific argumentation, allowing a variety of ideas to be presented, but insisting that challenge and negotiation also occur would indicate that group members were skeptical. Accepting ideas without accompanying reasons would result in a low score because it is a sign of credulous thinking. In other words, students must be willing to ask, “how do you know?” or “Are you sure?” Groups that respond to the ideas of others with comments such as “ok”, “that sounds good to me”, or “whatever you think is right” would score low on this item.

**Comments:**
5. The participants provided reasons when supporting or challenging an idea.  

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>

**Description:** Providing reasons to support or challenge a claim, conclusion, or explanation is a crucial characteristic of argumentation. Claims must have some support provided for them beyond simply restating the claim itself. Making claims with out support would result in a low score on this item and including any reason like “that’s what I think”, “it doesn’t make sense”, “the data suggests...” or “but that doesn’t fit with...” would result in a higher score. Note: Personal or past experiences count as a reason for this item.

**Comments:**

6. The participants based their decisions or ideas on inappropriate reasoning strategies.  

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>

**Description:** When people are trying to support ideas they often: (a) jump to hasty generalizations, (b) attribute causality to random events, (c) insist that a correlation is evidence of causality, and (d) exhibit a confirmation bias (for example saying, “now we need some data to prove this”). Groups that avoid inappropriate reasoning strategies or recognize them when they occur would score high on this item. Groups where these types of reasoning strategies are common would score low on this item.

**Comments:**

7. The participants attempted to evaluate the merits of each alternative claim or explanation in a systematic manner.  

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>

**Description:** This addresses the tentative or responsive nature of science. The idea that there is often more than one way to interpret data or evidence and that only through careful analysis can an idea be accepted or eliminated. This gets at the “gut” response factor. Conclusions are not based on opinion or inference.

**Comments:**

8. The participants relied on the “tools of rhetoric” to support or challenge ideas.  

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>

**Description:** “Tools of rhetoric” refer to tricks or strategies used to win a debate. Tool of rhetoric include: (a) claiming that if someone cannot disprove a claim it must be true, (b) using emotive words and false analogies, (c) directing the focus of the discussion from thinking about a claim or an explanation to thinking about the person holding or proposing a claim or an explanation, (d) over-relying on authorities, (e) dichotomizing issues so that if you discredit one position, then the observer is forced to accept the other view, and (f) making claims that are a simple restatement of one of the premises. Groups that avoided using the tools of rhetoric would score high on this item. Note: This item focuses on how the content of a discussion is presented or supported (i.e., how they are saying it) rather than the content of the discussion (i.e., what they are saying).

**Comments:**

---

**EPISTEMIC ASPECTS OF SCIENTIFIC ARGUMENTATION**

**HOW CONSISTENT THE PROCESS IS WITH THE CULTURE OF SCIENCE**

(These items target how the group determines what counts as valid or acceptable)
9. The participants used evidence to support and challenge ideas or to make sense of the phenomenon under investigation.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>

**Description:** A goal of scientific argumentation is the use of data as evidence to defend a claim, conclusion, or explanation. This item implies that students were attempting to use evidence in their arguments. This should more than an opinion; they must include data. Statements like “that’s what I think” or “it doesn’t make sense” would result in a low score. Statements like “the data we found suggests that...” or “our evidence indicates...” would result in a higher score.

**Comments:**

10. The participants examined the relevance, coherence, and sufficiency of the evidence.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>

**Description:** This item draws attention to the amount and kinds of evidence used to support a claim or explanation. Groups that attempt to (a) determine the value of a piece of evidences (e.g., “does that matter?”), (b) look at links or the relationship between multiple pieces of evidence (e.g., “This supports X and Y but this only supports X”), or (c) attempt to determine if there is enough evidence to support an idea (e.g., “We do not have any evidence to support that”) would score higher on this item.

**Comments:**

11. The participants evaluated how the available data was interpreted or the method used to gather the data.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>

**Description:** The evidence provided for a claim or explanation should be evaluated based on how well the data was gathered and interpreted. A question such as “Why is that evidence included?” or “How did they gather their data?” or “Where did that data come from?” indicates that the participants are assessing methods or an interpretation of data and would result in a higher score.

**Comments:**

12. The participants used scientific theories, laws, or models to support and challenge ideas or to help make sense of the phenomenon under investigation.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>

**Description:** Science is theory-laden. In other words, scientists rely on broad, well-supported organizing ideas to frame their arguments and claims. Students should also employ these paradigmatic ideas in providing warrants for the evidence and claims they make or use to refute others’ claims. Explicit reference to these “big ideas” will result in a higher score on this item.

**Comments:**
13. The participants made distinctions and connections between inferences and observations explicit to others.  

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>

**Description:** The structure of scientific arguments includes evidence involving both empirical (such as quantitative measurements and systematic observations) and inferential (noting of trends and logical connections among observations) aspects. Making these distinctions and their connections explicit to others enhances the quality of the argumentation and thus results in a higher score.

**Comments:**

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>

14. The participants used the language of science to communicate ideas.  

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>

**Description:** This item stresses the importance of the accurate use of scientific language by students. The adoption and use appropriate terms (e.g., condensation, force, etc), phrases (e.g., “it supports” rather than “it proves”) or ways of describing information is a characteristic of argumentation that is scientific. Note: Ideas may be explicated before being labeled with the correct terminology.

**Comments:**

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>

**SOCIAL ASPECTS OF SCIENTIFIC ARGUMENTATION**  
**HOW THE PARTICIPANTS INTERACT WITH EACH OTHER**  
(These items target group dynamics)  

15. The participants were reflective about what they know and how they know.  

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>

**Description:** It is important for members of the group to agree on what they know and to be specific about how they know. Statements such as, “do we all agree?” or “is there anything else we need to figure out?” or “can we be sure?” indicate that participants are monitoring their progress and have an end goal in mind.

**Comments:**

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>

16. The participants respected what each other had to say.  

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>

**Description:** Respecting what others have to say is more than listening politely or giving tacit agreement. Respect also indicates that what others had to say was actually heard and considered (e.g., “that is a good point”, interesting idea”, or “I hadn’t thought of that”). A group that scored high on this would allow everyone to present their ideas and express their opinions without censure or ridicule.

**Comments:**

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>

17. The participants discussed an idea when it was introduced into the conversation.  

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>
**Description:** To be a participating and contributing member of the group, it is important to feel valued. Ideas and opinions need to be critically acknowledged. This means they are considered and given weight by the group. Groups that ignore ideas when they are proposed (results in the same idea being mentioned over and over) would earn a low score on this item.

**Comments:**

<table>
<thead>
<tr>
<th>18. The participants encouraged or invited others to share or critique ideas.</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong> Good argumentation comes from considering and comparing competing ideas from multiple individuals to construct the most robust explanation of the phenomenon under study. Groups that consist of individuals that invite others to share (e.g., “what do you think”), critique (e.g., “do you agree” or “it is ok to disagree with me”), or discuss an idea (e.g., “let’s talk about this some more”) would score higher that a group with an alienating leader that dominates the conversation and the work of the group.</td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>

**Comments:**

<table>
<thead>
<tr>
<th>19. The participants restated or summarized comments and asked each other to clarify or elaborate on their comments.</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong> The depth of discussion will be enhanced by not making implicit judgments or assumptions about another person’s ideas or views, and it demonstrates that their point of view is valued and is furthering the discussion. Communication provides students with opportunities to identify the strengths and weaknesses of their understanding.</td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>

**Comments:**

<table>
<thead>
<tr>
<th>20. There was equal participation from all members of the group.</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong> The degree to which in member contributed to the argumentation impacts the depth and breadth of the discourse. Also, one or two high performers may result in a high score on some items, but not be representative of the actual argumentation event. Groups where some members are not engaged would score low on this item.</td>
<td>Not at all</td>
<td>Once or Twice</td>
<td>A few times</td>
<td>Often</td>
</tr>
</tbody>
</table>

**Comments:**

Total: /60
<table>
<thead>
<tr>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4</td>
</tr>
</tbody>
</table>

1. The author provided a well-articulated explanation that provides a **sufficient answer** to the research question.

| Not included. | The author provides a brief answer to the question that lacks detail. | The author provides an adequate answer to the question BUT it does not include everything that is needed. | The author’s explanation is detailed and includes everything that is needed AND the author expressed their ideas clearly and provided the reader with insight about the phenomenon. |

2. The author uses genuine evidence to support the explanation and **presents** the evidence in an appropriate manner.

| Not included. | The author does **NOT** present data to support their argument. | The author presents data to support their argument. BUT did **NOT** include correctly formatted diagram(s), graph(s) or table(s). OR **DID** NOT use correct units and significant figures. | The author uses data support their argument AND included correctly formatted diagram(s), graph(s) or table(s) AND correct units and significant figures. |

3. The author provides **enough evidence** to support the explanation AND the evidence is valid and reliable

| Not included. | The author **DID NOT** support all of their ideas with evidence OR used evidence based on unreliable or invalid data. | The author provides support for all of their ideas using valid and reliable data BUT used only some of the evidence to support each idea. | The author provides support for all of their ideas using valid and reliable data AND used all of the evidence to support each idea. |

4. The author’s **rationale** is sufficient and appropriate.

| Not included. | The author explains why the evidence was included OR why the evidence supports the explanation (but not both) BUT the rationale is **NOT** sound. | The author explains why the evidence was included OR why the evidence supports the explanation (but not both) AND the rationale is sound. | The author explains why the evidence was included AND why the evidence supports the explanation AND the rationale is sound. |

5. The author’s answer is **consistent** with what the scientific community accepts and/or with other groups in their lab section.

| The conclusion was inaccurate | There are major flaws in the conclusion and little or no comparison with other groups was made. | The conclusion is partially correct, but comparison with other groups was used to explain errors. | Results were compared in a meaningful way with other groups or with known values. The authors went beyond their own data in looking for an answer to the question. |
Office of the Vice President For Research
Human Subjects Committee
Tallahassee, Florida 32306-2742
(850) 644-8673, FAX (850) 644-4392

APPROVAL MEMORANDUM

Date: 2/18/2010

To: Joi Walker

Address: [REDACTED]
Dept.: MIDDLE AND SECONDARY EDUCATION

From: Thomas L. Jacobson, Chair

Re: Use of Human Subjects in Research
Fostering Science Proficiency in Undergraduate Chemistry Lab Sections

The application that you submitted to this office in regard to the use of human subjects in the proposal referenced above have been reviewed by the Secretary, the Chair, and two members of the Human Subjects Committee. Your project is determined to be Expedited per 45 CFR Â§ 46.110(7) and has been approved by an expedited review process.

The Human Subjects Committee has not evaluated your proposal for scientific merit, except to weigh the risk to the human participants and the aspects of the proposal related to potential risk and benefit. This approval does not replace any departmental or other approvals, which may be required.

If you submitted a proposed consent form with your application, the approved stamped consent form is attached to this approval notice. Only the stamped version of the consent form may be used in recruiting research subjects.

If the project has not been completed by 2/17/2011 you must request a renewal of approval for continuation of the project. As a courtesy, a renewal notice will be sent to you prior to your expiration date; however, it is your responsibility as the Principal Investigator to timely request renewal of your approval from the Committee.

You are advised that any change in protocol for this project must be reviewed and approved by the Committee prior to implementation of the proposed change in the protocol. A protocol change/amendment form is required to be submitted for approval by the Committee. In addition, federal regulations require that the Principal Investigator promptly report, in writing any unanticipated problems or adverse events involving risks to research subjects or others.

By copy of this memorandum, the Chair of your department and/or your major professor is reminded that he/she is responsible for being informed concerning research projects involving
human subjects in the department, and should review protocols as often as needed to insure that
the project is being conducted in compliance with our institution and with DHHS regulations.

This institution has an Assurance on file with the Office for Human Research Protection. The
Assurance Number is IRB00000446.

Cc: Victor Sampson, Advisor
HSC No. 2009.3756
FSU Behavioral Consent Form

Fostering Science Proficiency in Undergraduate Chemistry Lab Sections

You are invited to participate in a research study that is designed to examine the impact of a new instructional model called Argument-Driven Inquiry (ADI) that is designed to improve the science proficiency of undergraduate science students. You were selected as a possible participant because you are enrolled in an undergraduate science course that requires a laboratory-based experience. We ask that you read this form and ask any questions you may have before agreeing to be in the study.

Background Information

The overall goal of this study is to document how undergraduate students’ knowledge, skills, and attitudes toward science change over the course of a semester in response to the Argument-Driven Inquiry (ADI) instructional model. The research questions that will guide this study are:

- What is the overall effectiveness of the ADI instructional model on students understanding of scientific explanations and the nature of scientific knowledge, their ability to engage in scientific practices and discourse, and their interest in/attitude towards science?

Joi Phelps Walker is the principal investigator for this study. She is a Professor of Chemistry at Tallahassee Community College and a doctoral student in the School of Teacher Education at Florida State University.

Procedures

If you choose to participate in this study, you are giving me (and my research team) permission to use the work that you will turn in as part of this course as data for this study. The work that we will use includes your assessments (the Performance Tasks that that you will take at the beginning and end of the semester) and your lab reports. We will also videotape the interactive argumentation sessions that take place five times during the semester. This will require 15 to 20 minutes of video recording. These video recordings will not be seen by anyone outside the research team.

Your participation in this study will involve no additional work or time on your part other than what you already completed as part of this class. The information provided by you during this study will be used to evaluate the impact of the Argument-Driven Inquiry instructional model on undergraduate science learning.

Risks and Benefits of Being in the Study

There are no risks, or direct benefits, from your participation in this study. However, your participation in this study will enable us to improve undergraduate science education.

Confidentiality

The records of this study will be kept private and confidential to the extent allowed by law. In any sort of report we might publish, we will NOT include any information that will make it
possible to identify you. Research records will be stored securely and only members of my research team will have access to the records. We will keep the copies of your work in a locked file cabinet or password protected computer. These materials will be destroyed one year after the completion of this research project, which is June 2011.

**Voluntary Nature of the Study**

Participation in this study is voluntary. Your decision whether or not to participate will not affect your grade in the course or your current and future relations with Tallahassee Community College (TCC) or Florida State University (FSU). If you decide to participate, you are free to or withdraw at any time without affecting your grade or those relationships.

In addition, the small class size of this class or the fact that your instructor might be upset with you may make you feel like you have to participate in the study. You do not have to participate if you do not want to. In fact, your instructor will not know if you agreed to participate or not until after he or she submits your grade. Once again, your participation in this study is completely voluntary and choosing to or not to participate in this study will not influence your grade in any way, your instructor’s opinion of you, or your current or future relationship with TCC and FSU.

**Contacts and Questions**

You may ask any question you have now. If you have a question later, you are encouraged to contact Dr. Victor Sampson at:

School of Teacher Education  
205 Stone Building  
Florida State University

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher(s), you are encouraged to contact the FSU Institutional Review Board at 2010 Levy Street, Research Building B, Suite 276, Tallahassee, FL 32306-2742, or 850-644-8633, or by email at jjccoper@fsu.edu. You will be given a copy of this information to keep for your records.

**Statement of Consent**

I have read the above information. I have had the opportunity to ask questions and have the questions answered. I consent to participate in the study.

________________________________________
Print Name

________________________________________  ______________________________
Signature                                      Date
References


American Association for the Advancement of Science. (1989). *Project 2061: Science for all Americans*. Washington, DC.


presented at the Annual International Conference of the National Association of Research in Science Teaching: (NARST).


BIOGRAPHICAL SKETCH

Joi Phelps Walker received a Bachelor of Arts degree in chemistry from Mary Baldwin College in 1982. She received a Master of Science degree in organic chemistry from the University of Tennessee, Knoxville in 1986. Following her graduate work in chemistry she held positions as an educator, a research scientist, an environmental chemist and a forensic chemist. In 1996 she accepted a fulltime appointment to teach chemistry at Tallahassee Community College. In 2006, she entered the graduate program in Science Education at Florida State University which culminated in this document.