Argument-Driven Inquiry as a Way to Help Students Learn How to Participate in Scientific Argumentation and Craft Written Arguments: An Exploratory Study

VICTOR SAMPSON
School of Teacher Education and FSU-Teach, College of Education, The Florida State University, Tallahassee, FL 32306-4459, USA

JONATHON GROOMS
School of Teacher Education, College of Education, The Florida State University, Tallahassee, FL 32306-4459, USA

JOI PHELPS WALKER
Division of Science and Mathematics, Tallahassee Community College, Tallahassee, FL 32304, USA

Received 27 June 2009; revised 30 July 2010; accepted 26 August 2010

DOI 10.1002/sce.20421

ABSTRACT: This exploratory study examines how a series of laboratory activities designed using a new instructional model, called Argument-Driven Inquiry (ADI), influences the ways students participate in scientific argumentation and the quality of the scientific arguments they craft as part of this process. The two outcomes of interest were assessed with a performance task that required small groups of students to explain a discrepant event and then generate a scientific argument. Student performance on this task was compared before and after an 18-week intervention that included 15 ADI laboratory activities. The results of this study suggest that the students had better disciplinary engagement and produced better arguments after the intervention although some learning issues arose that seemed to hinder the students' overall improvement. The conclusions and implications of this research include several recommendations for improving the nature of laboratory-based instruction to help cultivate the knowledge and skills students need to participate in scientific argumentation and to craft written arguments. © 2010 Wiley Periodicals, Inc. *Sci Ed* **95:**217–257, 2011

Correspondence to: Victor Sampson; e-mail: vsampson@fsu.edu

INTRODUCTION AND OBJECTIVES

A major aim of science education in the United States is for all students to develop an understanding of scientific inquiry and the abilities needed to engage in this complex practice by the time they graduate from high school (American Association for the Advancement of Science, 1993; National Research Council [NRC], 1996, 2005, 2008). An important aspect of the process of scientific inquiry, which is often neglected inside the classroom, is argumentation. In science, argumentation is not a heated exchange between rivals that results in winners and losers or an effort to reach a mutually beneficial compromise; 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 an important practice in science that makes science different from other ways of knowing (Driver, Newton, & Osborne, 2000; Duschl & Osborne, 2002).

Duschl (2008) suggests that students need to develop several important and interrelated understandings and abilities to be able to participate in scientific argumentation (p. 277). First, an individual must be able to use important conceptual structures (e.g., scientific theories, models, and laws or unifying concepts) and cognitive processes when reasoning about a topic or a problem. Second, an individual must know and use the epistemic frameworks that characterize science to develop and evaluate claims. Third, and perhaps most importantly, individuals that are able to engage in scientific argumentation must understand and be able to participate in the social processes that shape how knowledge is communicated, represented, argued, and debated in science. Empirical research, however, indicates that most students do not develop this type of knowledge or these skills while in school because most students do not have an opportunity to engage in scientific argumentation or to learn how this practice differs from other types of argumentation (NRC, 2005, 2008; Duschl et al., 2007). One way to solve this problem, we argue, is to develop and encourage the use of new instructional models that change the nature of traditional laboratory activities so students have more opportunities to develop the understandings and abilities needed to participate in scientific argumentation over the course of an academic semester or year.

The overall goal of this study, therefore, was to explore how a new instructional model that we created to address this need influences the ways students participate in scientific argumentation and craft written arguments. This model, which we call Argument-Driven Inquiry or ADI, is intended to function as a template or a guide that science teachers can use to design laboratory activities that are more authentic (i.e., engages students in scientific practices such as argumentation) and educative (i.e., leads to better understanding and improved abilities) for students. To evaluate the promise and the potential of the model, we used a performance task to assess how six small groups of students participate in argumentation and craft written arguments before and after an 18-week intervention. The semester long intervention included 15 different laboratory activities that were designed using the ADI instructional model. We decided to focus on both process and product as dependent measures in this study to better reflect the multiple dimensions of this complex scientific practice and to help avoid biases that can result from only focusing on one outcome. However, in addition to the two desired outcomes, we also predicted that there might be several unintended or unanticipated learning issues (e.g., conceptual, cognitive, epistemic, or social) that might result from the use of this new instructional model in an authentic context such as a classroom. We, therefore, focused our analysis on the shortcomings or failures of the instructional model as well as the successes.

In light of the goals of the investigation outlined above and this analytical focus, the research questions that guided this study were as follows:

1. To what extent does a series of laboratory activities designed using the ADI instructional model influence the ways students participate in scientific argumentation and craft a written scientific argument?
2. Is there a relationship between the ways groups of students participate in scientific argumentation and the nature of the written arguments they create?
3. What types of learning issues need to be addressed to better help students learn how to participate in scientific argumentation and craft written scientific arguments?

THE ARGUMENT-DRIVEN INQUIRY INSTRUCTIONAL MODEL

A number of science education researchers (e.g., Driver et al., 2000; Duschl & Osborne, 2002; Duschl, 2008) have argued for the need to shift the nature of classroom instruction away from models that emphasize the transmission of ideas from teacher to students, to models that emphasize knowledge construction and validation though inquiry. As a result, a number of instructional models, such as the Science Writing Heuristic (Wallace, Hand, & Yang, 2005) and Modeling Instruction (Hestenes, 1992; Wells, Hestenes, & Swackhamer, 1995), have been developed in recent years to provide students with more opportunities to construct explanations that describe or explicate natural phenomena and to make them public by sharing them in small groups or in whole class discussions. These instructional models are designed to create a classroom community that will help students understand scientific explanations, learn how to generate scientific evidence, and reflect on the nature of scientific knowledge. The ADI instructional model is similar to these approaches because it is designed to change the nature of a traditional laboratory instruction so students have an opportunity to learn how to develop a method to generate data, to carry out an investigation, use data to answer a research question, write, and be more reflective as they work. The ADI instructional model, however, also provides an opportunity for students to participate in other important scientific practices such as scientific argumentation and peer review during a lab. It is through the combination of all these activities, we argue, that students can begin to develop the abilities needed to engage in scientific argumentation, understand how to craft written arguments, and learn important content as part of the process.

The current iteration of the ADI instructional model, which is a template or guide for designing a laboratory-based activity, consists of seven components or steps. We define the boundaries of the seven steps of the model by scope and purpose. Each step of the model, however, 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, we will describe each step and our rationale for including it in this instructional model. However, given our focus on cultivating the knowledge and abilities that students need to participate in scientific argumentation and to create high quality written arguments, we will devote more of our discussion to the stages that specifically target these outcomes.

The first step of the ADI instructional model is the *identification of the task* by the classroom teacher. In this step of the model the goal of the teacher is to introduce the major topic to be studied and to initiate the laboratory experience. This step is designed to capture the students' attention and interest. The teacher also needs to make connections between past and present learning experiences (i.e., what students already know and what they need to find out) and highlight the goal of the investigation during this step of the model. To accomplish this, we typically provide students with a handout that includes a brief introduction and a researchable question to answer, a problem to solve, or task to complete. The handout also includes a list of materials that can be used during the investigation and some hints or suggestions to help the students get started on the investigation. We also include information about what counts as a high quality argument in science and specific criteria that students can use to assess the merits of an argument in science that students can use as a reference during the third and fourth steps of the model.

The second step of the ADI instructional model is *the generation of data*. In this step of the model, students work in a collaborative group to develop and implement a method (e.g., an experiment, a systematic observation) to address the problem or to answer the research question posed during the first step of the model. The overall intent of this step is to provide students with an opportunity to learn how to design an informative investigation, to use appropriate data collection or analysis techniques, and to learn how to deal with the ambiguities of empirical work. This step of the model also gives students a chance to learn why some methods work better than others and how the method used during a scientific investigation is based on the nature of the research question, the phenomenon under investigation, and what has been done by others in the past.

The third step in the ADI instructional model is the *production of a tentative argument*. This component of the model calls for students to construct an argument that consists of a claim, their evidence, and their reasoning in a medium, such as a large whiteboard, that can be shared with others. In our research, we define a claim as a conclusion, conjecture, an explanation, or some other answer to a research question. The evidence component of an argument refers to measurements or observations that are used to support the validity or the legitimacy of the claim. 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 between variables. The reasoning component of an argument is a rationalization that indicates why the evidence supports the claim and why the evidence provided should count as evidence. Figure 1 provides a diagram that illustrates how we conceptualize these various components of a scientific argument.

This step of the model is designed to emphasize the importance of an argument (i.e., an attempt to establish or validate a claim on the basis of reasons) in science. In other words, students need to understand that scientific knowledge is not dogmatic and scientists must be able to support a claim with appropriate evidence and reasoning. It is also included to help students develop a basic understanding of what counts as an argument in science and how to determine whether the available evidence is valid, relevant, sufficient, and convincing enough to support a claim. More importantly, this step is designed to make students' ideas, evidence, and reasoning visible to each other; which, in turn, enables students to evaluate competing ideas and eliminate conjectures or conclusions that are inaccurate or do not fit with the available data in the next stage of the instructional model.

The fourth stage in the instructional model is an *argumentation session* where the small groups share their arguments with the other groups and critique the work of others to determine which claim is the most valid or acceptable (or try to refine a claim to make it more valid or acceptable). 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 other students, articulate more substantial warrants for their views, and evaluate the merits of competing ideas (Duschl et al., 2007; Linn & Eylon, 2006). In other words, argumentation sessions are designed to "create a need" (Kuhn & Reiser, 2006) for students to take a critical look at the product (i.e., claim or argument), process (i.e., method), and context (i.e., the theoretical foundation) of an inquiry. It also provides an

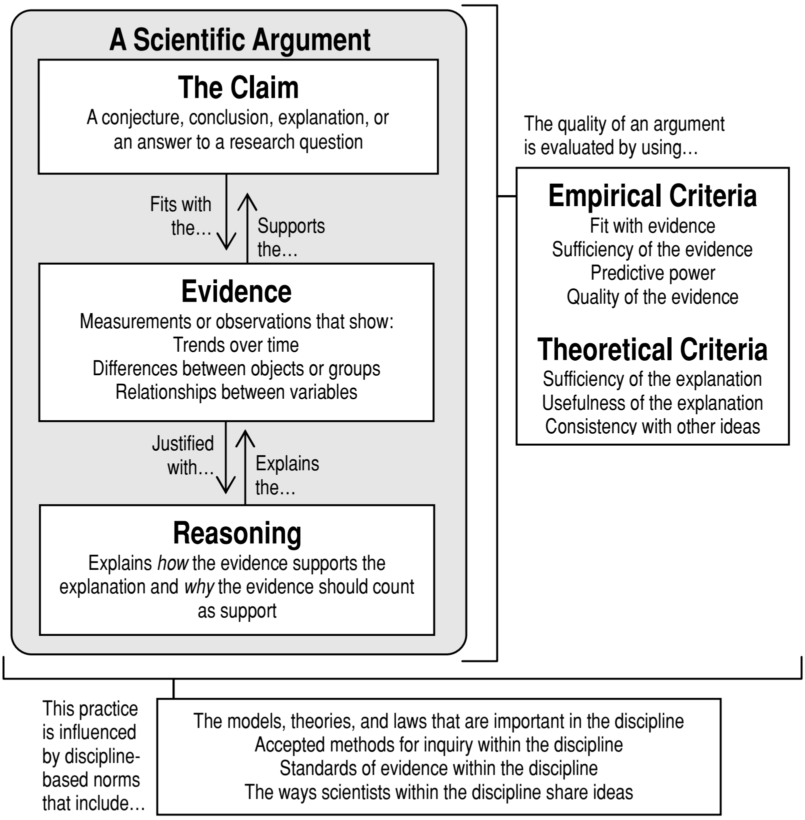

Figure 1. A framework that can be used to illustrate the components of a scientific argument and some criteria that can and should be used to evaluate the merits of a scientific argument.

authentic context for students to learn how to participate in the social aspects of scientific argumentation.

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 and by helping students negotiate and adopt new criteria for evaluating claims or arguments. This is important because current research indicates that students often have a repertoire of ideas about a given phenomenon “that are sound, contradictory, confused, idiosyncratic, arbitrary, and based on flimsy evidence” and that “most students lack criteria for distinguishing between these ideas” (Linn and Eylon, 2006, p. 8). Similarly, the work of Kuhn and Reiser (2005) and Sampson and Clark (2009a) suggests that students often rely on informal criteria, such as plausibility, the teacher’s authority, and fit with personal inferences, to determine which ideas to accept or reject during discussions and debates. We include the argumentation sessions as a way to help students learn how to use criteria valued in science, such as fit with evidence or consistency with scientific theories or laws, to distinguish between alternative ideas (see Figure 1 for other criteria that are made explicit to students). It also gives students an opportunity to refine and improve on their initial ideas, conclusions, or methods by encouraging them to negotiate meaning as a group (Hand et al., 2009). These sessions, in other words, are designed to encourage students to use the conceptual structures, cognitive processes, and epistemic frameworks of science to support, evaluate, and refine a claim.

The fifth stage of ADI is the *creation of a written investigation report* by individual students. We chose to integrate opportunities for students to write into this instructional model because writing is an important part of *doing* science. Scientists, for example, must be able to share the results of their own research through writing (Saul, 2004). Scientists must also be able to read and understand the writing of others as well as evaluate its worth. In order for students to be able to do this, they need to learn how to write in a manner that reflects the standards and norms of the scientific community (Shanahan, 2004). In addition to learning how to write in science, requiring students to write can also help students make sense of the topic and develop a better understanding of how to craft scientific arguments. This process often encourages metacognition and can improve student understanding of the content and scientific inquiry (Wallace, Hand, & Prain, 2004).

To encourage students to learn *how to write* in science and to *write to learn* about a topic under investigation, we use a nontraditional laboratory report format that is designed to be more persuasive than expository in nature. The format is intended to encourage students to think about what they know, how they know it, and why they believe it over alternatives. To do this, we require students to produce a manuscript that answers three basic questions: What were you trying to do and why?, What did you do and why?, and What is your argument? The responses to these questions are written as a two page "investigation report" that includes the data the students gathered and then analyzed during the second step of the model as evidence. Students are encouraged to organize this information into tables or graphs that they can embed into the text. The three questions are designed to target the same information that is included in more traditional laboratory reports but are intended to elicit student awareness of the audience, the multimodel and nonnarrative structure of scientific texts, and to help them understand the importance of argument in science as they write. This step of the model also requires each student to negotiate meaning as he or she writes and helps students refine or enhance their understanding of the material under investigation (Wallace et al., 2005; Hand et al., 2009).

The sixth stage of ADI is a *double-blind peer review* of these reports to ensure quality. Once students complete their investigation reports they submit three typed copies without any identifying information to 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 lab group along with a peer review sheet for each set of reports. The peer review sheet 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 are framed as 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?, and Is the author's reasoning sufficient and appropriate? The lab groups review each report as a team and then decide whether it can be accepted as is or whether it needs to be revised based on a *negotiated* decision that reflects 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 to improve the quality of the report and the writing as part of the review.

This 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 where students are expected to hold each other accountable. Students, as a result, should expect to discuss the validity or the acceptability of scientific claims and, over time, 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 scientific arguments (see Sampson, Walker, Dial, & Swanson, 2010, for more information about this process).

The seventh, and final, stage of the ADI instructional model is the *revision of the report* based on the results of the peer review. The reports that are accepted by the reviewers are given credit (complete) by the teacher and then returned to the author while the reports that need to be revised are returned to the author without credit (incomplete). These authors, however, are encouraged to rewrite their reports based on the reviewers' feedback. Once completed, the revised reports (along with the original version of the report and the peer review sheet) are then resubmitted to the classroom teacher for a second evaluation. If the revised report has reached an acceptable level of quality then the author is given full credit (complete). Yet, if the report is still unacceptable it is returned to the author once again for a second round of revisions. This step is intended to provide an opportunity for students to improve their writing mechanics, argument skills, and their understanding of the content without imposing a grade-related penalty. It also provides students with an opportunity to engage in the writing process (i.e., the construction, evaluation, revision, and eventual submission of a manuscript) in the context of science.

THEORETICAL AND EMPIRICAL FOUNDATION

Theoretical Perspectives Used to Develop ADI

The ADI instructional model is rooted in social constructivist theories of learning (see Driver, Asoko, Leach, Mortimer, & Scott, 1994; Anderson, 2007; Scott, Asoko, & Leach, 2007). This perspective suggests that learning science involves "people entering into a different way of thinking about and explaining the natural world; becoming socialized to a greater or lesser extent into the practices of the scientific community with its particular purposes, ways of seeing, and ways of supporting its knowledge claims" (Driver et al., 1994, p. 8). Thus, learning the practices of science such as scientific argumentation as well as the content of science (i.e., theories, laws, and models) involves both personal and social processes. The social process of learning involves being introduced to the concepts, language, representations, and the practices that makes science different from other ways of knowing. This process requires input and guidance about "what counts" from people that are familiar with the goals, norms, and epistemological commitments that define science as a community of practice. Thus, learning is dependent on supportive and educative interactions with other people. The individual process of learning, on the other hand, involves the construction of knowledge and understanding through the appropriation of important ideas, modes of communication, modes of thinking, and practices. This requires individuals to make sense of their experiences and the integration of new views with the old.

This theoretical perspective has two important consequences for instructional design and for what it means for students to learn science inside the classroom. First, students must engage in authentic scientific practices to learn from their experiences. Reiser and colleagues (2001) suggest that authentic scientific practices require students to engage in the reasoning and the discursive practices of scientists (such as coordinating theory and evidence to support an explanation) rather than the exact activities of professional scientists (such as grant writing or field work). Second, students must develop an understanding of what makes certain aspects of a practice more productive or useful than others and why. In science, for example, important practices include the ability to design and conduct informative investigations and to craft convincing arguments. However, what counts as "informative" and "convincing" in science reflect the epistemological commitments of the scientific community for what counts as scientific knowledge and what methods can be used to generate such knowledge (Sandoval & Reiser, 2004). These ideas make some practices in science (such as using empirical evidence to support a claim) more useful or important to scientists and makes science different from other ways of knowing. It is therefore important for students to understand what makes certain strategies or techniques more productive or useful to learn how to engage in authentic scientific practices in more productive ways. In other words, students' laboratory experiences need to also be educative in nature.

Given this theoretical perspective, the design of the ADI instructional model is based on the hypothesis that efforts to improve students' abilities to participate in scientific argumentation and to craft written arguments will require the development and use of laboratory activities that are more authentic and educative. In order for a laboratory activity to be more authentic, students need to have an opportunity to engage in specific practices that are valued by the scientific community (such as investigation design, argumentation, writing, and peer review). These types of authentic experiences, however, also need to be educative to promote student learning. To accomplish this requirement, mechanisms that enable students to not only see what they are doing wrong but also what they need to do to improve need to be built into each laboratory activity. This type of approach, where students have a chance to engage in authentic scientific practices and receive feedback about their performance, should enable learners to see why some techniques, strategies, tools, ways of interacting, or activities are more useful or productive than others in science as they complete the laboratory activities embedded into a course. It should also help students understand how scientific knowledge is developed and evaluated and how scientific explanations are used to solve problems. This approach, in turn, should enable students to develop more complex argumentation skills and a more fluid "grasp of practice" (Ford, 2008) that will enable them to use their knowledge and skills in different contexts or in novel situations.

How Students Participate in Scientific Argumentation and Craft Written Arguments

One of the main goals underlying the development of the ADI instructional model, as discussed in the Introduction of this article, is to provide teachers with a way to give students more opportunities to learn how to participate in scientific argumentation and to help them develop the knowledge and abilities needed to craft a written scientific argument during laboratory activities. This type of focus is important because current research indicates that students often struggle with the nuances of scientific argumentation despite being skillful at supporting their ideas, challenging, and counterchallenging a claim during conversations that focus on everyday issues (e.g., Baker, 1999; Pontecorvo & Girardet, 1993; Resnick, Salmon, Zeitz, Wathen, & Holowchak, 1993; Stein & Miller, 1993). This paradox, we argue, results from students' not understanding the goals and norms of scientific argumentation and how these goals and norms diverge from the forms of argumentation they are accustomed to rather than a lack of skill or mental capacity.

Students, for example, are often asked to gather data and then make sense of a phenomenon based on data when they engage in scientific argumentation inside the classroom. Research suggests that this aspect of scientific argumentation is often difficult for students. Students, for example, often do not seek out or generate data that can be used to test their ideas or to discriminate between competing explanations (e.g., Klahr, Dunbar, & Fay, 1990; Schauble, Klopfer, & Raghavan, 1991). In addition, students often use inappropriate or irrelevant data (McNeill & Krajcik, 2007; Sandoval & Millwood, 2005) or they only rely on their personal views to draw a conclusion (Hogan & Maglienti, 2001). Students also do not tend to use empirical and theoretical criteria valued by the scientific community to determine which ideas to accept, reject, or modify when they participate in scientific argumentation. Students, for example, often do not base their decisions to accept or reject an idea on the available evidence. Instead, students tend to use inappropriate reasoning strategies (Zeidler, 1997), rely on plausibility or fit with past experiences to evaluate the merits of an idea (Sampson & Clark, 2009a), and distort, trivialize, or ignore evidence in an effort to reaffirm their own conceptions (Clark & Sampson, 2006; Kuhn, 1989). These findings, however, should not be surprising given the few opportunities students have to gather and analyze data or evaluate ideas based on genuine evidence outside the science classroom.

Students also need to be able to generate explanations and craft a written argument that includes appropriate evidence and reasoning to participate in scientific argumentation. Current research indicates that these complex tasks are also difficult for students. For example, many students do not understand what counts as a good explanation in science (McNeill & Krajcik, 2007; Sandoval & Reiser, 2004; Tabak, Smith, Sandoval, & Reiser, 1996) so they tend to offer explanations that are insufficient and vague or they only offer a description of what they observed rather than providing an underlying causal mechanism for the phenomenon under investigation (Driver, Leach, Millar, & Scott, 1996; McNeill, Lizotte, Krajcik, & Marx, 2006; Sandoval & Millwood, 2005). Students also often find it difficult to differentiate between what is relevant and what is irrelevant data when crafting a written argument (McNeill & Krajcik, 2007) and often do not use sufficient evidence to support their claims (Sandoval & Millwood, 2005). Students also tend to rely on unsubstantiated inferences to support their ideas (Kuhn, 1991) or use inferences to replace evidence that is lacking or missing (Brem & Rips, 2000). Empirical research also indicates that students often do not provide warrants, or what some authors refer to as reasoning (e.g., Kuhn & Reiser, 2005; McNeill & Krajcik, 2007), to justify their use of evidence (Bell & Linn, 2000; Erduran, Simon, & Osborne, 2004; Jimenez-Aleixandre, Rodriguez, & Duschl, 2000). These observations, however, once again seem to reflect students' lack of understanding of the goals or norms of scientific argumentation and "what counts" in science rather than a unique mental ability.

To summarize, these studies indicate that students often struggle with many aspects of scientific argumentation in spite of their ability to support, evaluate, and challenge claims or viewpoints during everyday conversations. Students, in other words, seem to be able to participate in nonscientific forms of argumentation with ease, but often find it difficult to make sense of data, to generate appropriate explanations, and to justify or evaluate claims using criteria valued in science when they are asked to engage in more scientific forms of argumentation. Students also struggle to produce high-quality written arguments in science. Thus, the available literature indicates that secondary students have the cognitive abilities and social skills needed to participate in scientific argumentations, but need an opportunity to develop new conceptual, cognitive, and epistemic frameworks to guide their decisions and interactions in the context of science. We, therefore, developed the ADI instructional model as a way to help students learn the conceptual structures, cognitive processes, and epistemological commitments of science by giving them an opportunity to engage in scientific practices, such as investigation design, argumentation, and peer review, and making these important aspects of science explicit and valuable to the students.

METHOD

Although the literature reviewed here suggests that the ADI instructional model should be an effective way to help students learn how to participate in scientific argumentation and to produce high-quality written arguments, we decided to conduct an exploratory study to examine the potential and feasibility of the model as a first step in our research program. Our goal was to use the ADI instructional model to design a series of laboratory activities and then pilot them inside an actual classroom with one of the authors serving as the instructor of record. This type of study had several advantages given our research goals and questions. First, it allowed us to determine whether the model functions as intended in an actual, although in some ways atypical, classroom setting. Second, it allowed us to examine the changes in the ways students interacted with ideas, materials, and each other in greater detail than is often feasible in studies with larger samples. Finally, and perhaps most importantly at this stage of our research program, it permitted us to examine the successes and failures of the ADI instructional model so we can refine it to help improve student learning. This focus also enabled us to clarify our understanding of several dimensions of ADI that seem to contribute to changes in student practices and some learning issues that seem to arise when science educators attempt to make laboratory activities more authentic and educative for students.

Participants

Nineteen 10th-grade students (7 males, 12 females, average age = 15.4 years) chose to participate in this study. These students were all enrolled in the same section (23 students in total) of a chemistry course. The course was taught at a small private school located in the southwest United States that served families with middle to high socioeconomic status. The ethnic diversity of the student population at the school was 94.9% White and 5.1% African American. This school requires 4 years of science for graduation and follows a "physics first" science curriculum. This means that all the students enrolled at the school are required to take conceptual physics in 9th grade, chemistry in 10th grade, biology in 11th grade, and either advanced physics, chemistry, or biology in the 12th grade.

Procedure

The 19 participants were randomly assigned (by pulling names out of a jar) to one of six groups after the second day of class. Groups 1–5 were made up of three individuals, and Group 6 consisted of four individuals (due to the odd number of participants). Groups 1, 2, 3, and 5 each consisted of two females and one male, Group 4 consisted of three females, and Group 6 was three males and one female. Groups 1 and 3 each had a student who spoke Russian at home. Each group was then asked to complete a performance task (see the section Data Sources). The performance task required each group to make sense of a discrepant event and then generate a written argument that provided and justified the group's explanation. All six groups completed this task during a lunch period or after school prior to the first ADI lab investigation without any input or support from the classroom teacher. Each group worked in an empty room and in front of a video camera so that the interactions that took place between the students and the available materials could be recorded. At the conclusion of the 18-week intervention (see the section The Intervention), the six original groups were asked to complete the same performance task for a second time. As before, each group completed the task during a lunch period or after school without any input or support from the classroom teacher and each group worked in an empty room in front of a video camera.

We chose to use the same performance task as a pre- and postintervention assessment in this study to facilitate comparisons over time. Given the substantial literature that indicates that the nature of argumentation that take place within a group is influenced by a wide range of contextual factors (such as object of the discussion, the available resources) and not just the argumentation skills of the participants (Andriessen, Baker, & Suthers, 2003), we needed to ensure that the complexity of the task, the underlying content, and the materials available for the students to use were the same during both administrations of the assessment. It is important to note, however, that the use of an identical assessment pre- and postintervention can result in a testing effect in some situations.

The testing effect refers to the robust finding that the act of taking a test not only assesses what people know about a topic but also tends to lead to more learning and increased long-term retention of the material that is being assessed (Roediger & Karpicke, 2006). There are several factors that can contribute to a testing effect (see Roediger & Karpicke, 2006, for an overview); however, the two most serious issues are (1) when a test provides additional exposure to the material (i.e., overlearning) and (2) when individuals are able to learn from their mistakes during the first administration of a test (i.e., feedback). We, as a result, attempted to limit these two potential sources of error by using an assessment that required the students to generate an original and complex explanation for an ill-defined problem rather than having them select from a list of several options (see the section Data Sources). We also did not give the students any feedback about their performance after the first administration of the assessment. It is important to acknowledge, however, that the students in this study might have continued to think about the problem after the first classroom experience with it, which may have artificially inflated the overall quality of the arguments crafted by each group postintervention. This issue, unfortunately, could not be controlled for given the nature of the research design employed and is therefore a limitation of this study.

The Intervention

All the students enrolled in the chemistry course participated in 15 different laboratory activities that were designed using the ADI instructional model. Table 1 includes an overview of each ADI laboratory activity. All 15 of these investigations included the seven stages of the ADI instructional model that were outlined earlier. For each lab, the students worked in a collaborative team of three or four. Students were randomly assigned to a new team after each lab so that all the students had an opportunity to work in a wide variety of groups throughout the 18-week semester.

There were four types of ADI investigations (see Table 1). The goal of the first type of investigation was to develop a new explanation. In these investigations, students were asked to explore a phenomenon (such as the macroscopic behavior of matter) and then to create an explanation or model for that phenomenon. This type of investigation was used as a way to introduce students to an important theory, law, or concept in science (such as the molecular-kinetic theory of matter) and was the focus of six different labs. The goal of the second type of investigation was to revise an explanation. In these investigations, students were asked to refine and expand on an explanation they developed in a previous investigation so they could use it to explain a different but related phenomenon. This type of investigation was the focus of two different labs. The goal of the third type of investigation was to evaluate an explanation. In these investigations, students were provided with a scientific explanation (such as the law of conservation of mass) or several alternative explanations and then asked to develop a way to test it or them. This type of investigation was the focus of two different labs. The goal of the fourth, and final, type of investigation was to use an explanation to solve a problem. In these investigations, students were asked to use a concept introduced in class (such as molar mass or types of chemical reactions) to solve a problem (identify an unknown powder or the products of a reaction). This type of investigation was the focus of five different labs.

TABLE 1
Overview of the 15 Argument-Driven Inquiry Laboratory Activities

Lab	Type of Investigation	Overview of the Laboratory Activity
1	Develop a new explanation	The students are introduced to Aristotle's model of matter and asked to develop a better explanation for the behavior of matter based on data they collect about the behavior of gases, liquids, and solids when heated and when matter is mixed with other forms of matter.
2	Revise an explanation	The students are asked to revise the explanation they developed during Lab #1 so they can also use it to explain the difference between heat and temperature. To do this, students collect data about the rate of diffusion of a gas at different temperatures and temperature changes in water when it s heated and/or mixed with water at a different temperature.
3	Revise an explanation	The students are asked to revise their model from lab #2 so they can also use it to explain what happens to matter at the submicroscopic level during a chemical reaction. To do this, students collect data about six different chemical reactions and two different physical changes.
4	Evaluate an explanation	The students develop and implement a method to test the validity of the law of conservation of mass.
5	Develop a new explanation	The students develop an explanation for the structure of the atom based on 14 observations about the characteristics of atoms gathered through empirical research.
6	Use a scientific explanation to solve a problem	The students develop and implement a method to identify six different compounds using the atomic spectra of 10 known compounds.
7	Develop a new explanation	The students develop a way to organize 30 elements into a table based on similarities and differences in their physical and chemical properties that will allow them to predict the characteristics of an unknown element.
8	Use a scientific explanation to solve a problem	The students develop and implement a method to determine whether density is a periodic property or not using elements from group 4A.
9	Develop a new explanation	The students develop a principle to explain why specific elements tend to form one type of ion and not another based on the characteristics of 21 different elements.
10	Develop a new explanation	The students develop and implement a method to identify factors that affect the rate at which an ionic compound dissolves in water. Then the students develop an explanation to for why the factors they identified influence the rate at which a solute dissolves in water.
11	Develop a new explanation	The students investigate the solubility of ionic, polar, and nonpolar compounds in a variety of polar and nonpolar solvents. The students create a principle to explain their observations.

(*Continued*)

TABLE 1
Continued

Lab	Type of Investigation	Overview of the Laboratory Activity
12	Use a scientific explanation to solve a problem	The students are given seven containers filled with seven "unknown" powders. The students must identify each unknown from a list of 10 known compounds based on the concept of molar mass.
13	Use a scientific explanation to solve a problem	Students are given two different unidentified hydrates. The students then develop and implement a method to identify these hydrates from a list of possible unknowns based on the concept of chemical composition.
14	Use a scientific explanation to solve a problem	Students determine the products produced in six different chemical reactions based on the concepts of solubility, polyatomic ions, and common types of reactions (i.e., synthesis, decomposition, single replacement, double replacement, and combustion).
15	Evaluate an explanation	Students are provided with three alternative chemical reactions for the thermal decomposition of sodium chlorate. The students then develop and implement a method to determine which chemical equation is the most valid or acceptable explanation.

The students also participated in a variety of activities that were designed to introduce or reinforce important content before or after each laboratory experience during the intervention (see Table 2). These activities included, but were not limited to, listening to short targeted lectures (L), partaking in whole class discussions (WCD), engaging in group work (GW), completing practice problems (PP), watching demonstrations (D), and completing readings (R) selected from the course textbook (Suchocki, 2000). These activities reflect "commonplace" teaching practices that are often observed in high school science classrooms (Stigler, Gonzales, Kawanaka, Knoll, & Serrano, 1999; Weiss, Banilower, McMahon, & Smith, 2001). We predicted, however, that these "commonplace" teaching activities would do little to influence the ways student participate in scientific argumentation or how they craft written arguments given the available literature. Table 2 provides an overview of the classroom activities by day and the amount of time spent on each activity for the entire 18-week intervention.

The students also completed a number of assessments throughout the 18-week semester in addition to the laboratory experiences and other classroom activities. The classroom teacher used these instruments for both formative (FA) and summative assessment (SA) purposes (see Table 2). These instruments, however, were not deemed suitable for research purposes. Therefore, any information about student learning or understanding that was collected by the instructor using these instruments was not included as a source of data in this study.

Data Sources

We used a performance task, as noted earlier, to assess how the students participate in scientific argumentation and craft a scientific argument. This performance task, which we call the *candle and the inverted flask* problem (see Lawson, 1999, 2002), required the small groups of students to negotiate a shared understanding of a natural phenomenon and then

TABLE 2
Classroom Activities By Day Over the Course of the Intervention

Week	Day				
	Monday	Tuesday	Wednesday	Thursday	Friday
1	L	GW (PreIPT)	L – PP (PreIPT)	CA – R (PreIPT)	ADI Lab #1
2	ADI Lab #1 cont. (Molecular Kinetic Theory of Matter A)				L – D
3	No School	ADI Lab #2 (Molecular Kinetic Theory of Matter B)			
4	L – PP – GW	ADI Lab #3	No School	No School	ADI Lab #3 cont.
5	ADI #3 cont. (Molecular Kinetic Theory of Matter C)			L – D	No School
6	L – PP	ADI Lab #4	No School	ADI Lab #4 cont. (Conservation of Matter)	
7	WCD – PP	CA	ADI Lab #5 (Structure of Atom A)		
8	ADI Lab #5 cont.		L – D	ADI Lab #6 (Structure of the Atom B)	
9	ADI Lab #6 cont.	L – PP	ADI Lab #7 (Periodic Trends A)		
10	L – D	ADI Lab #8 (Periodic Trends B)			No School
11	L – PP	ADI Lab #9 (Periodic Trends C)			WCD – PP
12	CA	ADI Lab #10 (Solubility A)			
13	L – D	R – PP – GW	ADI Lab #11 (Solubility B)		
14	L – D	L – PP	ADI Lab #12 (Chemical Composition A)		
15	R – PP	ADI Lab #13 (Chemical Composition B)			
16	WCD – PP	L – D – PP	ADI Lab #14 (Chemical Reactions A)		
17	ADI #14 cont.		R – PP – GW	ADI Lab #15 (Chemical Reactions B)	
18	ADI #15 cont.		L – R (PostIPT)	WCD – PP (PostIPT)	CA (PostIPT)

Note: PreIPT = PreIntervention Performance Task completed by the groups after school or during a lunch period (two/day), R = Reading from the textbook, L = Lecture, GW = group work, PP = practice problems, WCD = Whole class discussion, D = Demonstration, CA = Classroom assessment of student learning, PostIPT = Post-Intervention Performance Task completed by the groups before school, after school, or during a lunch period (two/day).

develop a written scientific argument that provides and justifies an explanation for it. The problem begins with a burning candle held upright in a pan of water with a small piece of clay. A flask is then inverted over the burning candle and placed in the water. After a few seconds, the candle flame goes out and water rises in the flask. Students are then asked: Why does the water rush up into the inverted flask? Students are given a pan of water, a flask, a graduated cylinder, five candles, a book of matches, a stopwatch, a wax pencil, and a ruler and then directed to use these materials to generate the data they will need to answer the research question. Once the group develops and agrees upon a sufficient answer, they are required to produce a written argument that provides and justifies their conclusion with evidence and reasoning.

The students needed to explain two observations to provide a sufficient answer to the research question posed in this problem. First, they needed to explain why the flame goes out. Second, they had to explain why the water rises into the flask. The generally accepted explanation for the first observation is that the flame converts the oxygen in the flask to carbon dioxide until too little oxygen remains to sustain combustion. The generally accepted explanation for the second observation is that the flame transfers kinetic energy to gas molecules inside the flask. The greater kinetic energy causes the gas to expand and some of this gas escapes out from underneath the flask. When the flame goes out, the remaining molecules transfer some of their kinetic energy to the flask walls and then to the surrounding air and water. This transfer causes a decrease in gas pressure inside the flask. The water inside the flask then rises into the flask until the air pressure pushing on the outside water surface is equal to the air pressure pushing on the inside surface (Birk & Lawson, 1999; Lawson, 1999; Peckham, 1993).

A common student explanation for these observations is the idea that oxygen is "used up." The loss of oxygen results in a partial vacuum inside the flask. Water is then "sucked" into the flask because of this vacuum. Most students, however, fail to realize that when oxygen "burns" it combines with carbon (i.e., combustion) to produce an equal volume of CO_2 gas inside the flask (Lawson, 1999). Students also often fail to realize that a vacuum cannot "suck" anything. Rather the force causing the water to rise is a push from the relatively greater number of air molecules hitting the water surface outside the flask (see Lawson, 1999, for a more detailed description of this phenomenon and for additional examples of student alternative conceptions).

These complex and ill-defined problems provided us with a unique context to examine how students participate in scientific argumentation and craft a written scientific argument with the same task. The counterintuitive and collaborative nature of the problem required the students to propose, support, challenge, and refine ideas to establish or validate an explanation. These discussions provided us with a way to observe how these students participated in scientific argumentation. The final arguments that the groups created during the task also supplied us with useful information about how these students' articulate and justify explanations. We choose to use the same task before and after the intervention, as noted earlier, to facilitate comparisons because the nature of argumentation is context dependent and is therefore influenced by more than just the skills of the participants. We also wanted the students to attempt to explain a phenomenon that was not studied in class but could be explained using content introduced during the course (e.g., the molecular-kinetic theory of matter, the conservation of mass, the difference between heat and temperature).

Data Analysis

Our main interest, given the goal and research questions of this study, was to document any changes in the two outcomes measures and to explore how the various components of the ADI may have supported the development of new ways of thinking and behaviors. To do this, we transcribed the videotapes of the conversations that took place within each group during the candle and the inverted flask problem. The transcription focused specifically on the sequence of turns and the nature of the interactions rather than speaker intonation or other discourse properties. Transcripts were parsed into turns, which were defined as segments of speaker-continuous speech. If an interruption stopped the speaker from speaking, the turn was considered complete, even if the content of the turn was resumed later in the conversation. If the student did not stop talking even though someone else was speaking, then all of the content was considered to be part of that same turn. One-word utterances, such as "yeah," "uhm," and so on, were also considered to be turns.

Coding schemes were then developed to document any potential changes in the ways students participated in scientific argumentation and to score the quality of the written arguments produced by each group before and after the intervention. Two researchers then used these coding schemes to independently evaluate the transcripts and the answer sheets. To assess the interrater reliability of the various coding schemes, a portion of the codes generated by each researcher for each outcome measure was compared. Cohen's κ values ranged from a low of 0.72 to a high of 0.90. Although a Cohen's κ value of 0.7 or greater indicates strong interrater reliability (Fleiss, 1981), all discrepancies between the two researchers were discussed and definitive codes were assigned once the two researchers reached consensus. The data presented in the Results section reflects these definitive codes.

Assessing How Students' Participate in Scientific Argumentation. To examine changes in the ways these students participated in the process of scientific argumentation, our analysis focused on the ways the students interacted with each other, ideas, and the available materials. Our analysis, however, went beyond simply documenting how often an individual student contributed to the activity or the conversation. Instead, we looked for evidence that the students participated in argumentation in a manner that was grounded in the disciplinary norms of science. To do this, we used Engle and Conant's (2002) definition of disciplinary engagement as an analytical framework.

Engle and Conant (2002) define engagement in terms of students actively speaking, listening, responding, and working as well as high levels of on-task behavior. Greater engagement can be inferred when more students in the group make substantive contributions to the topic under discussion, and their contributions are made in coordination with each other. Engagement also means that students attend to each other and discuss ideas when other students propose them. Finally, it means that few students are involved in few unrelated or off-task activities. Very few off-task comments were observed in any of the groups before or after the intervention. Therefore, to examine any changes in the students level of engagement, we first used a coding scheme developed by Barron (2000, 2003) to examine how students responded to the various ideas (both content and process related) introduced into the discussion.

The intent of this analysis was to determine how the students reacted to ideas. Four categories of responses were used: *accept*, *discuss*, *reject*, and *ignore*. Accept responses included any reaction where an individual voiced agreement with the speaker, supported the proposal, or incorporated the idea into the group's argument. Discuss responses included any reaction that resulted in further discussion of the idea. Examples of this type of response include questioning the rationale behind an idea, challenging it with new information or a different idea, asking for clarification, and revising or adding to an idea. Reject responses included any reaction that voiced disagreement with the speaker or made a claim that an idea was incorrect, irrelevant, or not helpful to the task at hand. Finally, ignore responses were coded as not giving a verbal response to an idea when it was proposed. Definitions and examples for each code are provided in Table 3 (Cohen's $\kappa = 0.79$).

We then developed a coding scheme to capture the overall nature or function of the contributions the students made to the conversation when discussing the merits of an idea. We conducted this analysis as a second measure of engagement to determine how often the group members were questioning or challenging each other's ideas. This type of interaction, as suggested by Osborne, Erduran, and Simon (2004) is an important aspect of argumentation in general, because "episodes [of argumentation] without rebuttals have the potential to continue forever with no change of mind or evaluation of the quality of the substance of an argument" (p. 1008). The comments made during these episodes were coded using four categories: *information seeking*, *expositional*, *oppositional*, and *supportive*. The unit of analysis for these codes was conversational turns within a discuss episode. The start and end point of a discuss episode was defined as the first comment after an introduced idea and the first comment that indicated a new topic of discussion. Table 4 provides more detail about this coding approach (Cohen's $\kappa = 0.77$).

These hallmarks of engagement, although important, do not ensure that the students are interacting with others or ideas in a manner that reflects the discipline of science. Therefore, the notion of disciplinary engagement expands the concept of engagement to include scientific content and the goals and the norms or epistemological commitments of science. To determine whether the student engagement in scientific argumentation was disciplinary or not, we used two criteria. First, we examined how often the students used rigorous criteria valued in science, such as how well the idea fits with available evidence or

TABLE 3
Codes Used to Examine the Ways Group Members Respond to Proposed Ideas

Code	Definition	Examples
Accept	Any response where an individual voices agreement with the speaker, supports the proposal, or incorporates the idea into the group product but does not result in further discussion.	"Yeah, that makes sense" "You're right" "Let's write that down"
Reject	Any response that voices disagreement with the speaker or makes a claim that an idea is incorrect and the response does not result in further discussion.	"That's not it" "That can't be right"
Discuss	Any response that results in further discussion of an idea. Examples of this type of response include questioning the rationale behind an idea, challenging it with new information or a different idea, asking for clarification, and revising or adding to an idea.	"What do mean by that?" "Are you sure?" "But why does the water rise when the candle goes out?" "What if we say. . . "
Ignore	Not giving a verbal response to an idea when it was proposed.	

how consistent the idea is with accepted theories, models, and laws (Passmore & Stewart, 2002; Stewart, Cartier, & Passmore, 2005), to support or challenge an idea, conclusion, or other claim (i.e., an aspect of the epistemic framework that makes science different from other ways of knowing; see Duschl, 2008). Then we decided to examine how often the students used a scientific explanation (e.g., theories, models, and laws) when talking and reasoning about the phenomenon under investigation (i.e., the conceptual structures and cognitive processes of science; see Duschl, 2008).

Our first step in this part of the analysis was to develop a coding scheme to document the nature of the criteria the students were using to either justify or refute their ideas. Two categories of criteria were used: *rigorous* and *informal* (Cohen's $\kappa = 0.73$). Rigorous criteria include the reasons or standards that reflect the evaluative component of the argument framework outlined in Figure 1. Examples of rigorous criteria include fit with data (e.g., "but the water went higher in the flask with two candles"), sufficiency of data (e.g., "you do not have any evidence to support that"), coherence of an explanation (e.g., "how can something use up and produce oxygen at the same time?"), adequacy of an explanation (e.g., "that doesn't answer the question"), and consistency with scientific theories or laws (e.g., "but the law of conservation of mass says matter cannot be destroyed"). Informal criteria include reasons or standards that are often used in everyday contexts but are less powerful for judging the validity of an idea in science. Examples of informal criteria include appeals to authority (e.g., "well that's what she said"), discrediting the speaker (e.g., "he never knows what to do"), plausibility (e.g., "that makes sense to me"), appeals to analogies (e.g., "this is just like fits with personal experience (e.g., "that happened to me once"), judgments about the importance of an idea (e.g., "that doesn't matter"), and consistency with personal inferences (e.g., "candles use up oxygen so there must be a vacuum inside the flask").

We then developed a coding scheme to describe the nature of the content-related ideas that were spoken aloud during the discussion (Cohen's $\kappa = 0.75$). To do this, we used Hunt

TABLE 4
Codes Used to Examine the Overall Nature and Function of the Contributions During Discuss Episodes

Discourse Move	Definition	Examples
Information seeking	Comments used by an individual to gather more information from others. These utterances include requests for (a) additional information about the topic, (b) partners to share their views, (c) partners to clarify a preceding comment, or (d) information about the task.	"What did you mean by that?" "What do you think?" "Why?"
Expositional	Comments used by an individual to (a) articulate an idea or a position, (b) clarify a speaker's own idea or argument in response to another participant's comment, (c) expand on one's own idea, or (d) support one's own idea.	"I think the candle uses up all the oxygen" "I mean. . . "
Oppositional	Comments used by an individual to (a) disagree with another, (b) disagree and offer an alternative, (c) disagree and provide a critique, or (d) make another support his/her idea.	"That can't be right" "How do you know it used up all the oxygen?"
Supportive	Comments used by an individual to (a) elaborate on someone else's ideas, (b) indicate agreement with someone else's ideas, (c) paraphrase someone else's preceding utterance with or without further elaboration, (d) indicate that one has abandoned or changed an idea, (e) combines ideas, separates one idea into two distinct ideas, or modify an idea in some way, (f) justify someone else's idea or viewpoint, or (g) steer or organize the discussion or how people are participating in the discussion.	"Right" "That is just what I was thinking" "You're right, I was wrong" "That is just like. . . "

and Minstrell's (1994) and Minstrell (2000) facet analysis approach to examine the content of the students' comments. Facets are ideas that lack the structure of a full explanation and can consist of nominal and committed facts, intuitive conceptions, narratives, *p*-prims, or mental models based on experiences at various stages of development and sophistication (Clark, 2006). Examples of content-related ideas that were identified in this analysis include inaccurate facets of student thinking such as "there is nothing inside the flask," "the vacuum sucks the water up," and "the flame creates a vacuum" and accurate facets such as "oxygen is transformed into carbon dioxide" and "gas expands as it heats up." We then specifically looked to see whether the students mentioned the scientific explanations introduced in class during these discussions. The four scientific explanations that were introduced in class that were needed to develop an accurate explanation for the candle and the inverted flask problem, as noted earlier, were the kinetic-molecular theory of matter, the conservation of mass, the process of combustion, and the gas laws.

Assessing the Written Scientific Arguments. To examine changes in the students' ability to craft a scientific argument before and after the intervention, we examined the overall quality of the written arguments produced by each group. We focused on four specific aspects of a written argument that are often used in the literature to assess quality (see Sampson & Clark, 2008, for a review of this literature). The four aspects were (a) the adequacy of the explanation, (b) the conceptual quality of the explanation, (c) the quality of the evidence, and (d) the sufficiency of the reasoning. Each aspect was given a score based on the presence or absence of specific components on a four-point (0–3) scale. The scores earned on each aspect were then combined to assign an overall score to an argument. As a result, argument scores could range from 0 to 12, with higher scores representing a higher quality argument.

We assessed the adequacy of the explanation in the arguments by evaluating how well the explanation answered the research question (Cohen's $\kappa = 0.81$). An adequate explanation, given the problem posed in this study, needed to (a) explain why the candle went out, (b) explain why water rose into the inverted flask, and (c) provide an account for how these two observations were related. Arguments that included an explanation with more of these components, regardless of their accuracy from a scientific perspective, were scored higher on this aspect of quality than arguments that included an explanation that contained only one or two of these components.

The conceptual quality of the explanation (Cohen's $\kappa = 0.85$) was assessed using the same facet analysis approach (Hunt & Minstrell, 1994; Minstrell, 2000) outlined earlier. The various facets found within an explanation were identified and coded as accurate (e.g., heat causes the gas to expand, the flame converts oxygen to carbon dioxide) or inaccurate (e.g., the fire destroys all the oxygen in the flask, the vacuum sucks the water up). Explanations that contained more accurate ideas received higher scores than explanations composed of inaccurate ideas or explanations that contained a mixture of accurate and inaccurate ideas.

To assess the quality of the evidence, we examined whether or not appropriate and relevant evidence was used to support the given explanation (Cohen's $\kappa = 0.72$). Appropriate evidence was defined as measurements or observations that were used to demonstrate (a) a difference between objects or groups, (b) a trend over time, or (c) a relationship between two variables. Inappropriate evidence included (a) unjustified inferences, (b) appeals to hypothetical examples, (c) appeals to past instances or experiences, and (d) appeals to authority figures. Evidence was scored as relevant if it directly supported an aspect of the explanation. Arguments that included more appropriate and relevant evidence were scored higher than arguments that contained inappropriate but relevant evidence or appropriate but irrelevant evidence.

We then assessed the sufficiency of the reasoning by determining how well the group linked the evidence to the explanation and justified the choice of evidence (Cohen's $\kappa = 0.76$). Sufficient reasoning was defined as (a) an explicit explanation for how the evidence supports components of the explanation and (b) an explicit explanation of why the evidence should count as evidence. The presence of these two elements was then used to score the overall quality of the reasoning. Arguments that included reasoning in this manner thus received a higher score than arguments that provided evidence without justification or provided only simple assertions such as "it proves it" or "it just makes sense" as a way to link the evidence to the explanation.

Examining the Relationship Between the Process and Product of Scientific Argumentation. To determine whether there was a relationship between the level of a group's disciplinary engagement in scientific argumentation and the quality of the written arguments they developed in this context, we needed to first calculate a composite argumentation score.

We relied on three aspects of our previous analysis to accomplish this task. The first two aspects were the proportion of discuss responses to a proposed idea (see Table 3) and proportion of oppositional comments made during a discuss episode (see Table 4). These aspects were included in the composite score because they provided a measure of group engagement. The third aspect was how often the individuals within a group used rigorous criteria valued in science to evaluate or justify ideas. We included this aspect in the composite score because it provided a measure of the disciplinary nature of the argumentation that took place within the groups.

To calculate the composite scores, we first rank ordered the observed proportions of a specific type of comment within each aspect regardless of time (12 values per aspect). We then assigned a score of one to the bottom quartile of values, a score of two to the next quartile of values, and so on for each aspect. Finally, we summed the scores a group earned on the four different aspects of argumentation before the intervention to create an overall preintervention composite score and all the scores earned by a group after the intervention to create a postintervention score. The composite argumentation scores for each group can range from a low of 4 points to a high of 12 points (with higher scores representing greater disciplinary engagement). We then compared the argumentation composite score to the written argument score of each group both pre- and postintervention.

RESULTS

The presentation of our results is organized around the two main outcomes of interest and the relationship between the two. In each subsection, we will provide descriptive and inferential statistics to help illustrate the differences we observed in the performance of the groups. We will also provide representative quotations from the transcripts and the written arguments crafted by some of the groups to help support our assertions and to illustrate patterns and trends in the practices of these students.

The Students' Disciplinary Engagement in Scientific Argumentation

We first examined how often group members contributed to the discussion as the first measure of engagement in scientific argumentation. Figure 2 provides the number of total comments and the proportion of comments made by each student in all six groups before and after the intervention. These data indicate that the level of participation in four of the six

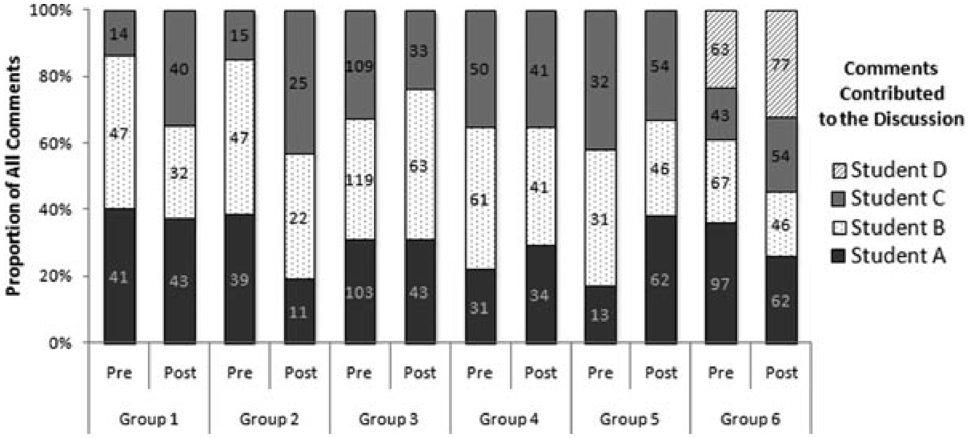

Figure 2. The number and proportion of comments contributed by each group member pre- and postintervention. Note: Groups 1–5 consisted of three students, and Group 6 consisted of four students.

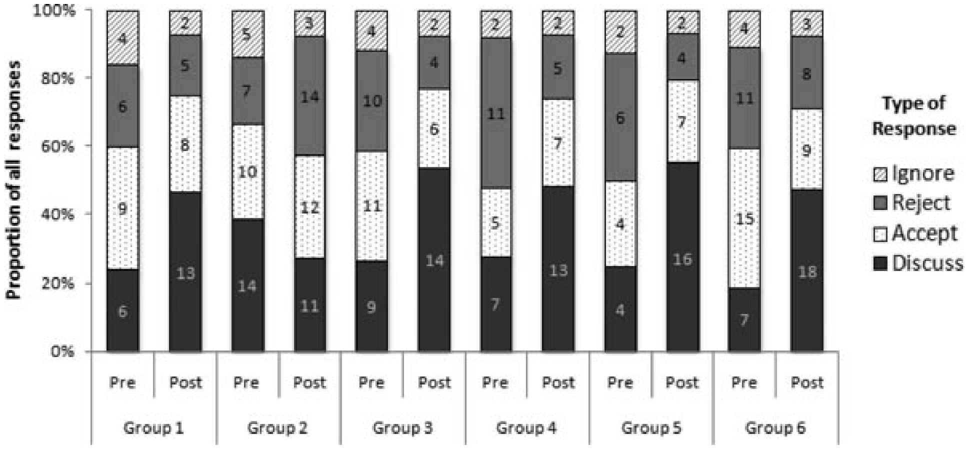

Figure 3. How group members responded to an idea when it was introduced into the conversation pre- and postintervention.

groups was much more balanced after participating in the 15 different ADI lab experiences. This pattern is well illustrated by the students assigned to Group 1. This group was one of the most lopsided in terms of participation at the beginning of the semester. The individual that made the fewest contributions to the conversation in this group (student 1-C) made 14% of the comments whereas the other two students made 46% (student 1-B) and 40% (student 1-A) of the total contributions. At end of the semester, however, the student that made the fewest comments in this group (student 1-B) made 28% of the contributions to the conversation and the other two made 37% (student 1-A) and 35% (student 1-C) of the comments. This represents a substantial shift in the levels of engagement by individual students and a much better balance of participation. This pattern held true for Groups 4, 5, and 6 as well.

We also, as noted earlier, examined how often group members discussed an idea when it was proposed as a second measure of engagement in scientific argumentation. Figure 3 provides the number and proportion of the four different types of responses (i.e., discuss, accept, reject, and ignore) in the six groups before and after the intervention. As this figure shows, all the groups with the exception of one (Group 2) had a lower a proportion of ignore, reject, and accept responses and a higher proportion of discuss responses after the intervention. A chi-square goodness-of-fit test confirmed that the observed pre- and postintervention differences were statistically significant, $\chi^2(3) = 14.52$, $p = .002$. These results indicate that more students in these groups were making substantive contributions to the discussion after the intervention. To illustrate this change, consider the following examples.

In the first example, taken from Group 3 before the intervention began, the various group members propose a number of ideas, but these ideas are rejected, accepted, or ignored without discussion.

Student 3-B: I already know what it is guys. It's suffocating the candle.
Student 3-C: No, no that's not it.
Student 3-A: What about the smoke?
Student 3-B: All the oxygen is being used up.
Student 3-C: Yeah, that sounds right.
Student 3-A: I still think it is the smoke.
Student 3-B: That's not it.

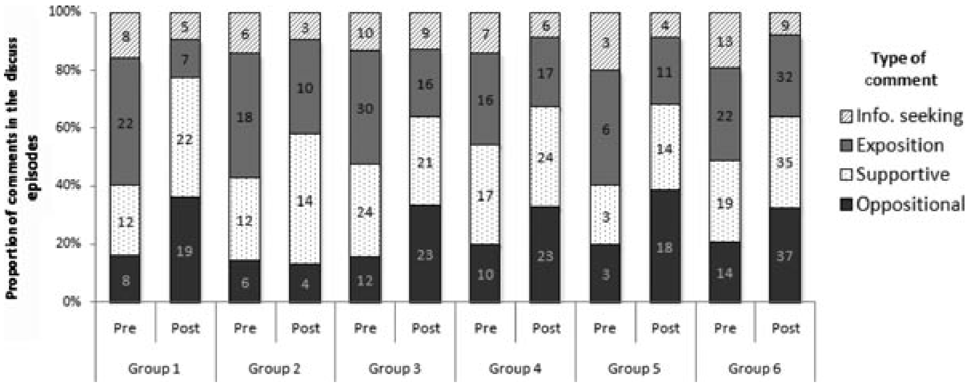

Figure 4. Types of comments group members made during discuss episodes pre- and postintervention.

These types of "that's not it" reject responses and "yeah, that sounds right" accept responses were common in the dialogue that took place within the groups prior to the intervention. As a result, these students rarely examined the underlying reasons for or against a particular idea or explanation. The groups instead seemed to spend a majority of their time indicating that they were either for or against a particular idea.

In contrast, when an idea was proposed after the intervention, it often served as a starting point for a more in-depth discussion. This trend is well illustrated in the following example. This excerpt is once again taken from Group 3 to help illustrate this change in the way the groups engaged in argumentation.

Student 3-B: When it . . . so this goes into here, and burns it up, creating smoke and it goes out. Then, the water's forced to go up.
Student 3-C: Why do you think that makes the water go up?
Student 3-B: Well, yeah . . . um, 'cause of the loss of oxygen. It's basically sucking it up into the thing because the oxygen is gone.
Student 3-A: But won't the smoke take up the space of the oxygen?
Student 3-C: Yeah, there's no way for smoke to come out because of the glass. [touches flask]
Student 3-B: Yeah. That makes sense, so what do you think it is?

Unlike the previous example, the students did not accept or reject the initial explanation outright. Instead the response of student 3-C led to a more in-depth discussion of the core issues involved in the problem (why the candle goes out and why the water rises into the flask). The greater frequency of discuss responses after the intervention indicates that the students were more engaged and were more willing to talk about, evaluate, and revise ideas. This type of interaction is important because some of the potential benefits of engaging in scientific argumentation with others seem to be lost when groups reject or accept ideas without discussing them first (Sampson & Clark, 2009a). Overall, this analysis suggests that these students were better able or more willing to engage in argumentation after participating in the 15 laboratory experiences designed using the ADI instructional model.

These data also suggest that these students were challenging each other's ideas and claims more frequently after the intervention. Figure 4 provides the proportion of information seeking, exposition, oppositional, and supportive comments during the discuss episodes before and after the intervention. As shown in Figure 4, most of the comments made by students during the discuss episodes before the intervention were devoted to exposition (i.e., proposing, clarifying, or justifying one's own idea) or were supportive (i.e., summarizing, revising, justifying, or adding to the ideas of others) and only a small proportion of the comments was oppositional in nature (i.e., simple disagreements and disagreements accompanied by critiques). This trend, however, did not continue after the intervention. In all the groups, except for one (Group 2), there were a much greater proportion of oppositional comments during the discuss episodes. A chi-square goodness-of-fit test confirmed that this observed difference was statistically significant as well, $\chi^2(3) = 31.21$, $p < .001$. Overall, these results indicate that these students were more skeptical, or at least more critical, of ideas after the intervention.

To illustrate this trend, consider the following examples. In the first example there are two discuss episodes. These discuss episodes are representative of the overall nature of the discourse that took place between individuals when discussing the merits of an idea before the intervention. The conversation in this example includes comments that focus on exposition or were supportive in nature. In other words, during these episodes the students are clarifying and justifying their own idea or revising, justifying, and adding to the ideas of the other members of the group.

Student 1-A: When the oxygen is removed from the air, the pressure. . .
Student 1-B: Inside the glass increases. Then why. . . ? I guess it's taking out the air so it's. . . you know.
Student 1-A: And the water is drawn to it.
Student 1-B: Alright. That makes sense.
Student 1-A: Cuz it's you know, it's . . . like a suction cup or something.
Student 1-C: Yeah. [End of discuss episode 1]
Student 1-B: Oh, so I think the clay also has something to do with it, cuz it's almost like a stopper.
Student 1-A: Thank you. Wait, what do you mean a stopper?
Student 1-B: It's just, it's just making, like if there wasn't the clay, obviously the candle wouldn't stay up. . .
Student 1-A: Yeah, it would go out if the clay wasn't holding it up. I mean you need to have the clay there.
Student 1-B: Totally. [End of discuss episode 2]

This excerpt is representative of the overall nature of the discussion that took place within the groups when group members did not accept or reject an idea outright before the intervention. During these episodes, there were few instances where students actually challenged an idea. Instead, the students in these groups spent the vast majority of their time either elaborating on an idea and asking questions or agreeing with and supporting the ideas of the other group members. For example, in the first discuss episode, rather than attempting to challenge the accuracy of an erroneous idea proposed by student 1-B (the pressure inside the flask increases) or requiring student 1-B to justify this idea, student 1-A simply added to the idea ("the water is drawn to it"). In the second discuss episode, student 1-A simply asks for clarification ("what do you mean a stopper?") when an idea ("the clay also has something to do with it") is proposed and then elaborates on student 1-B's idea ("Yeah, it would go out if the clay wasn't holding it up"). These types of interaction were common before the intervention. The students seemed unwilling to disagree, challenge, or critique the ideas of other group members (even when an idea that was introduced into the discussion was inaccurate from a scientific perspective).

Now compare the above example with the following excerpt of dialogue taken from the same group after the intervention. In this second example, the discourse is more oppositional in nature.

Student 1-C: So . . . the water is not causing the candle goes out.
Student 1-B: Why do you say that? I mean . . . like, how do you know it is not the water?
Student 1-C: It doesn't look like it's the water putting out the candle because the candle went out before the water ever actually touched the flame.
Student 1-A: Are you sure? Why don't we try it and check.
Student 1-C: Ok. [Student 1-C puts flask down over a lit candle.] Now watch.
Student 1-A: You're right. It went out before the water touched the wick. [End of discuss episode 1]
Student 1-C: Why don't we try doing this with an unlighted candle. That way we can see if it is the fire that is causing the water to rise . . . even though I think it's pretty safe to assume that the water won't do anything if it's unlit.
Student 1-B: Yeah I don't think it will. [Student 1-C puts flask down with candle unlit.]
Student 1-A: Nope.
Student 1-C: Yeah, so it's definitely the fire that causes it to rise—something the fire is doing, and the only thing that the fire is doing inside the flask is it's consuming the oxygen because there's really nothing else for it to consume. For the fire to burn away the wick, there has to be oxygen to react with. So, when the oxygen has been used up in there, we've got the partial vacuum in there.
Student 1-B: But how do you know it used up all the oxygen?
Student 1-C: Why else would the candle go out? [End of discuss episode 2]

This excerpt is representative of a substantial number of exchanges that took place after the intervention during the discuss episodes. These data suggest that the students were much more willing to disagree, challenge, or critique ideas when others proposed them. Furthermore, this type of oppositional discourse did not lead to the polarization of viewpoints or cause group members to opt out of the discussion. Instead, this type of discourse appeared to play an important role in moving the discussion forward and helped lead to the co-construction of a shared explanation. These disagreements and critiques, as suggested by Osborne et al. (2004) and Sampson and Clark (2009a), often led to a critical examination of an idea or the evidence and reasoning supporting a claim. Overall, this analysis indicates that the students seemed to be much more comfortable with oppositional discourse, which is an important characteristic of better argumentation in general, after the intervention.

These data also indicate that these students were engaging in argumentation in a manner that better reflects the discipline of science after the intervention. Students, in general, seemed to adopt and use more rigorous criteria to distinguish between explanations and to justify or evaluate ideas as a result of the intervention. Figure 5 shows how often (as a percentage of the total number of instances) individuals in the groups used rigorous and informal criteria to support or evaluate a claim, explanation, or other idea before and after the intervention. As illustrated in Figure 5, the data indicate that the groups as a whole relied on rigorous criteria 43% of the time and informal criteria 57% of the time before the intervention. At the end of the semester, in contrast, the groups as a whole used rigorous criteria 74% of the time and informal criteria 26% of the time. In addition, five of the six groups relied on rigorous criteria more frequently after the intervention. The one group (Group 2) that did not make a substantial gain in this regard, however, was already using rigorous criteria at a higher level than most of the other groups at the beginning of the semester. The results of a chi-square goodness-of-fit test confirmed that these observed differences were statistically significant, $\chi^2(1) = 47.78$, $p < .001$. These observations

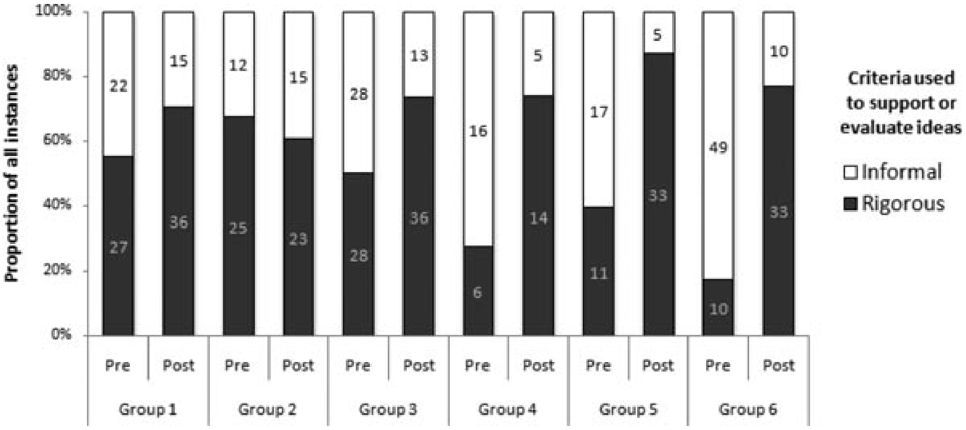

Figure 5. Types of criteria students used to support or challenge ideas when engaged in argumentation.

suggest that these students learned the criteria that we emphasized for evaluating the merits of explanations and arguments and then adopted them as their own as a result of the intervention.

To illustrate this trend, consider the following examples. In the first example, the students from Group 3 were relying on plausibility, personal inferences, and past experiences to evaluate the merits of an idea once it was introduced into the discussion. In other words, these students judged the validity or acceptability of ideas by how well they fit with their personal viewpoints.

Student 3-A: Ok, first of all guys. It's not asking "What is making the fire go out?" It is asking "why does the water rush up into the inverted flask?"
Student 3-C: Because it's the suction, like . . . it's like suction. Like when you suck on a straw.
Student 3-A: That sounds good to me.
Student 3-B: No it's not suction. That means that there would have to be an opening right here, and something would . . . something like a vacuum cleaner would have to suck the air out. That's the only way to get suction.
Student 3-A: Ok . . . how about this then. I think that since the candle's warm it causes smoke and the smoke causes the water rise . . .
Student 3-B: That doesn't make any sense.

Comments such as "that sounds good to me" and "that doesn't make any sense" were common in the discussion before the intervention. The high frequencies of these types of comments suggest that these students did not rely on rigorous criteria that are valued in science, such as fit with data, to evaluate, or support ideas before participating in the ADI lab experiences. After the intervention, however, these same students were more likely to use rigorous criteria when supporting and critiquing ideas. In this example, the students in Group 3 are attempting to evaluate the validity or acceptability of an idea by assessing how well the idea fits with their observations.

Student 3-C: Watch, when I hold the flask over the candle, it's going to keep going, but when I put it down, it goes out. [Student 3-C sets the flask over the candle and lets go. Candle goes out.] So it won't let oxygen in and the candle uses up the oxygen.

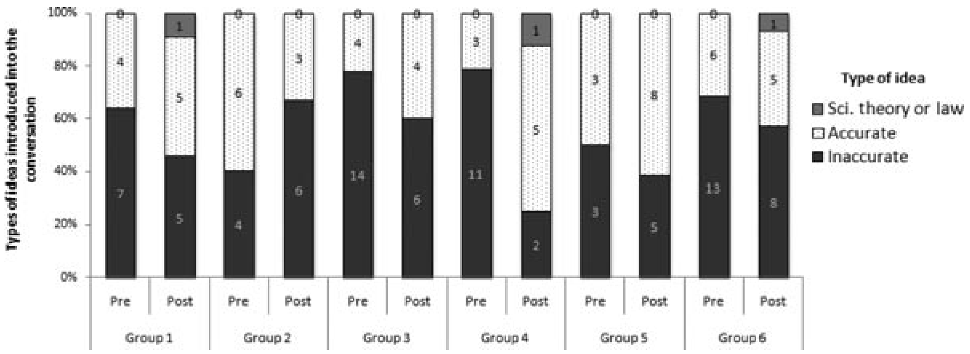

Figure 6. The number and proportion of inaccurate, accurate, and scientific theories or laws that were mentioned over the course of the performance task.

Student 3-B: Since there's no oxygen, it's trying to get to the oxygen on top, right? How is that possible?
Student 3-A: Because the oxygen is coming in with the water.
Student 3-C: But if that was true, why didn't the water keep going up?
Student 3-A: Because you let go.
Student 3-C: Oh.
Student 3-B: So keep on holding on. Try holding on.
Student 3-C: Ok. [Student 3-C lights the candle and puts the flask over the candle in water but not all the way down. Candle goes out]
Student 3-B: Ok, so what does that tell us. It needs oxygen, so the water is being forced into an isolated area with no oxygen. Is there any air being forced out? What do you think? Would there be any air being forced out?
Student 3-A: I don't know. How could we test that?

This excerpt is representative of many of the exchanges that took place within the groups after the intervention. Students in these groups seemed to rely on more rigorous criteria to distinguish between competing conjectures or ideas as they worked. These students, for example, would often generate an idea and then use the available materials as a way to test its merits. Although the students still used fit with a personal viewpoint as a criterion some of the time, the individuals in these groups used criteria that are more aligned with those valued in science with greater frequency after the intervention. This suggests that these students adopted and used new standards to evaluate or validate knowledge in the context of science.

The students in this study, however, did not use the conceptual structures of science (i.e., important theories, laws, or concepts) much when attempting to make sense of their observations before or after the intervention. Figure 6 provides the number and proportion of inaccurate ideas (e.g., there is a vacuum inside the flask), accurate ideas (e.g., the pressure is less inside the flask), scientific theories or laws (e.g., the conversation of mass) that were mentioned by at least one group member over the course of the conversation. As illustrated in Figure 6, no one in any of the groups mentioned a scientific theory or law before the intervention. After the intervention, there was not much difference; three of the six groups did not mention a single scientific explanation and the other three groups only mentioned one (the kinetic-molecular theory of matter in Groups 1 and 4 and the gas laws in Group 6). These results indicate that the students did not use scientific theories or laws to make sense of their observations or to critique the merits of a potential explanation before or after the intervention.

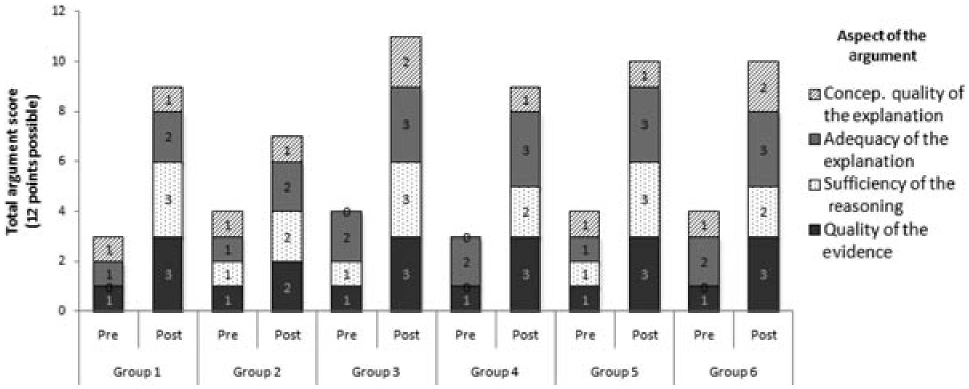

Figure 7. The overall score and the score on each aspect of the written argument produced by each group before and after the intervention.

These results, on the other hand, do indicate that all the groups but one (Group 2) mentioned a greater proportion of accurate ideas overall after the intervention, $\chi^2(1) = 4.45$, $p = .03$. This observation suggests that the students' understanding of the relevant content, as a whole, was better at the end of the intervention even though the students did not make explicit references to the scientific theories or laws discussed in class as they worked. This finding, however, was not unexpected given the length of the intervention, the number of laboratory activities, and the instructional activities that took place between each lab.

The Students' Ability to Craft a Written Scientific Argument

Figure 7 provides a comparison of the overall quality of the written argument each group produced before and after the intervention. The average score of the written arguments before the intervention was 3.6 (out of 12 possible points). The average score after the intervention, in contrast, was 9.3. This represents a 158% increase in the average overall quality of the written arguments produced by these groups. A Wilcoxon Signed-Rank test confirmed that the observed difference in the scores pre- and postintervention was statistically significant despite the small sample, $z = -2.26$, $p = .02$, two tailed. This analysis suggests that these students were able to craft higher quality arguments (in terms of the adequacy and conceptual quality of the explanation, the quality of the evidence, and the sufficiency of the reasoning) after participating in 15 ADI lab experiences.

The aspects of the students written arguments that showed the greatest improvement as a result of the intervention was the quality of the evidence and the sufficiency of the reasoning. To illustrate this trend consider the following written arguments. The first example, which was created by Group 4 at the beginning of the semester, exemplifies the type of argument crafted by the students[1] before the intervention:

> *What is your explanation?* Oxygen was taken away so the fire went out. The water was then sucked into the flask because a partial vacuum was created.

---

[1] Each group crafted an argument on the answer sheet by responding to three prompts: What is your explanation?, What is your evidence?, and What is your reasoning? The prompts were included on the answer sheet to help increase the reliability of the coding schemes. This is important because students often use inferences as evidence, which often makes it difficult for researchers to differentiate between the various components of student-generated arguments. See Erduran et al (2004) and Sampson and Clark (2008) for a discussion of this issue.

> *What is your evidence?* In a vacuum, there is less pressure, therefore, there is nothing holding the water down. The air pressure pushing down is less than water pressure pushing up.
>
> *What is your reasoning?* Because the flame consumed all the oxygen inside the bottle, it then had no fuel, and went out. This created a vacuum and caused the water to rise.

This argument included an explanation that provided a reason for the candle going out and a reason for the water rising into the flask but did not connect these two aspects of their explanation (2 out of 3 points). This explanation, however, contained only inaccurate facets so the conceptual quality was scored as poor (0 out of 3 points). The group then used an inference as evidence to support their conclusion. Although this is not appropriate evidence given our theoretical framework, it was scored as low (1 out of 3 points) because it was relevant to the provided explanation. Finally, the sufficiency of the reasoning was scored as poor (0 out of 3 points) because the group simply rephrased their initial explanation and did not explain why the evidence supports the explanation or why they choose to use that type of evidence. After the intervention, however, Group 4 produced a better argument:

> *What is your explanation?* The flame consumes the oxygen inside the flask and creates a partial vacuum. This lowers the air pressure inside the flask. The water is then pushed into the flask because the air pressure outside the flask is greater than it is inside the flask.
>
> *What is your evidence?* When we used two candles, the water went up more than it did with only one candle. It also takes one candle longer to go out (6.8 seconds) than it takes two candles to go out (4.5 seconds).
>
> *What is your reasoning?* The flame needs oxygen to fuel it. Once the oxygen is consumed the flame disappears. As the amount of oxygen decreases inside the flask so does the air pressure. Our data indicates that this process happens quicker when more candles are used because more candles consume the oxygen in a shorter amount of time.

This argument, unlike the groups' first attempt, includes an explanation that provides a reason for the candle going out, a reason for the water rising into the flask, and an explicit connection between these two aspects of the explanation (3 out of 3 points). However, this explanation contains a mixture of accurate (water is pushed into the flask, the air pressure outside the flask is greater) and inaccurate facets (creates a partial vacuum, etc.) so the conceptual quality of the entire explanation was scored as low (1 out of 3 points). The group then included two pieces of appropriate and relevant evidence to support the explanation (3 out of 3 points). The students' reasoning explains why the evidence supports the explanation but does not justify their choice of evidence (2 out of 3 points). Overall, this argument is a good representation of the nature of the written arguments produced by the six groups after the intervention. The arguments, in general, included a more adequate explanation and better evidence and reasoning, but the explanation was often conceptually inaccurate.

This improvement in the quality of the written arguments seemed to be due, in large part, to the students' lack of familiarity with the nature of scientific arguments at the beginning of the semester rather than a lack of skill or natural ability. This lack of familiarity with scientific arguments often resulted in students not understanding what counts as evidence and reasoning or what makes evidence different from reasoning. To illustrate this confusion, consider the following excerpt that shows how students in Group 4 talked about the evidence and reasoning components of an argument before the intervention. In this example, the students have decided on their explanation (i.e., the answer to the research question) and are in the process of crafting their argument.

Student 4-C: Wait, no, there's one more question . . . What is your reasoning?
Student 4-A: Don't look at me . . . I don't know.
Student 4-C: I don't understand . . . I don't understand the difference between evidence and reasoning.
Student 4-B: Yeah, I don't either.
Student 4-C: So how am I supposed to make the answer for reasoning different from the answer we already wrote?
Student 4-A: Just summarize it or write the explanation again.

This excerpt is representative of many of the exchanges that took place between students as they worked to develop their written argument prior to the intervention. Students clearly did not understand what counts as evidence and reasoning in the context of science. As a result, the arguments the students crafted often included an inference or a single observation as evidence and a simple restatement of the groups' explanation for reasoning. After the intervention, however, the students seemed to have a much better understanding of the nature of scientific arguments due to the explicit focus on the nature and structure of arguments in science. To illustrate this difference, consider the following excerpt (from Group 4 postintervention):

Student 4-A: Ok . . . now we need to give our evidence.
Student 4-C: All the variables we tested, all the things we measured, we should use that as our evidence.
Student 4-A: Anything else?
Student 4-B: We need to include our reasoning.
Student 4-A: Didn't we kind of do that already?
Student 4-B: No, the reasoning is why.
Student 4-C: They're not both why.
Student 4-B: I know. The evidence is our observations and the reasoning is why the observations support our explanation.

This excerpt is representative of many of the exchanges that took place within the groups after the intervention. These exchanges suggest that the students developed a better understanding of what counts as an explanation, evidence, and reasoning in a scientific argument. Although these students still struggled to produce an explanation that was accurate from a scientific perspective, the overall quality of the written arguments improved pre- to postintervention. These observations, when taken together, indicate that these students developed a more nuanced understanding of the various components of a scientific argument (based on how we defined them in our framework) and learned how to craft a better scientific argument over the course of the semester by participating in the 15 ADI lab experiences.

The Relationship Between the Process and Product of Argumentation

Figure 8 shows the relationship between the argumentation composite scores and the written argument scores earned by each group pre- and postintervention. As illustrated in this scatter plot, groups that had higher composite argumentation scores also had higher written argument scores, $r(12) = .89$, $p < .01$. The results of this analysis also indicate that all of the groups had low argumentation and argument scores before the intervention

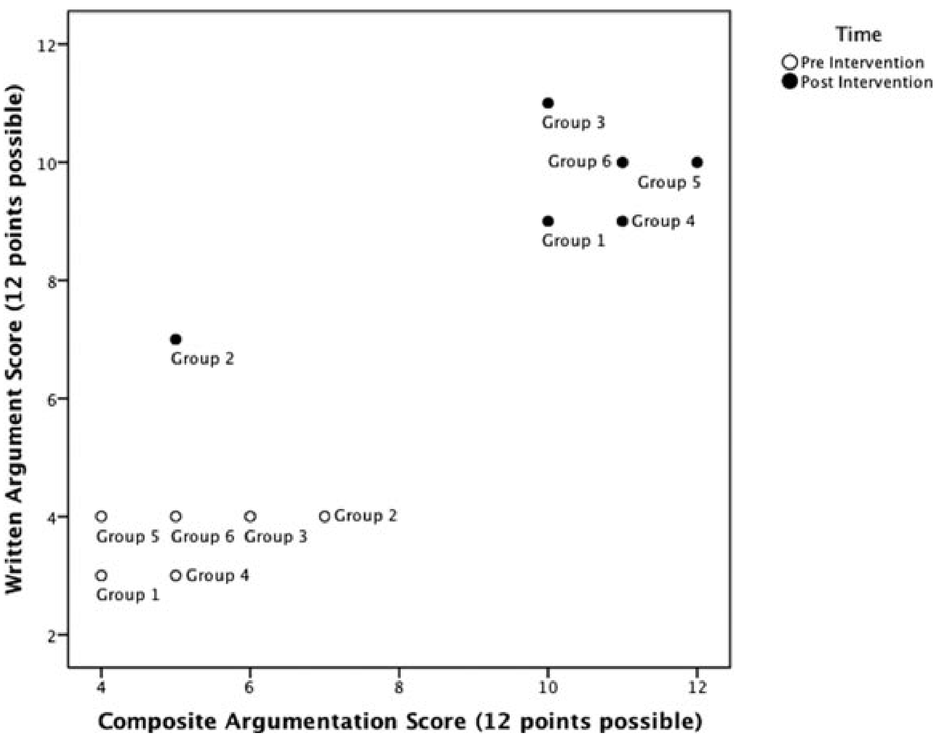

Figure 8. The relationship between the product and process of argumentation pre- and postintervention.

and all of the groups, with the exception of Group 2, had higher scores postintervention. These observations, while keeping in mind the small sample size, suggest that there is a relationship between the level of disciplinary engagement in argumentation and the overall quality of the written arguments crafted by these groups.

LIMITATIONS AND CONCLUSIONS

One of the overall goals for our current research program is to develop an instructional model that teachers can use to help students develop the understandings and abilities needed to participate in scientific argumentation and to craft written arguments during laboratory activities. In this article, we have chosen to focus on the theoretical and empirical framework underlying the design of the ADI instructional model and to present some of the data that have helped us to refine our understanding of the learning issues at hand. Our goal in this study has not been to evaluate the effectiveness of the ADI instructional model in comparison to other instructional approaches. If that were the goal, our method would have been poorly suited to the task. Instead, our goal has been to understand whether the ADI instructional model appears to function in a classroom setting in a manner predicted by the available literature and our theoretical framework. We were also interested in learning more about the ways students engage in argumentation and the nature of the arguments they create when they have more opportunities to construct an explanation and evaluate claims, evidence, and reasoning during laboratory activities. This study, therefore, should be viewed as a formative exploration rather than as a summative evaluation.

It is important to note, however, that understanding how a specific instructional model, such as ADI, functions inside a classroom is difficult. This is due, in large part, to the nature of the instructional model used, the complex nature of an existing setting, the nature of the content, and how these three factors (among others) interact with each other. Therefore, rather than attempting to isolate the effects of the various components of the model, we felt it was important to examine the impact of the model as a whole because the pieces of complex instructional models are dependent on each other and often have no effect in isolation (see Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003; Salomon, 1993). This may lead someone to wonder, however, about how we can tell whether or not the instructional model is worth refining and how to refine it. We believe that we can answer these two questions by looking at how the students' behavior or ways of thinking changed pre- to postintervention and then identifying potential underlying reasons for a change or a lack of change. Once we understand which behaviors or way of thinking did or did not change (or changed in an unpredicted or undesired manner) and the factors that seem to influence this process, we can then redesign one or more components of the model to address the observed shortcomings. We can then implement the revised model in a similar context and determine whether the shortcomings have been corrected or not.

Yet, despite this focus, our conclusions still need to be viewed in light of three main limitations. First, this study was designed as a way to explore "what could be" in science classrooms if the nature, focus, and the number of laboratory activities used over the course of a semester changed quite a bit from the status quo. This type of study, as Schofield (1990) suggests, requires a researcher to identify or create a context that is "ideal or exceptional on some a priori basis" (p. 217) in order "to see what is actually going on in there" (p. 217). Our findings, therefore, might be atypical due to the unique nature of the intervention. Second, we did not include a comparison group in this study to help control for the influence of time or a testing effect. Therefore, we can only speculate about how much change we would have observed in the ways these students participated in scientific argumentation and the overall quality of the arguments they created in response to a series of traditional laboratory activities or with no intervention at all. We felt, however, that the inclusion of a comparison group was not essential in this exploratory study given the substantial amount of literature that indicates that students do not learn how to participate in scientific argumentation or how to craft scientific arguments while in school (Duschl et al., 2007; Duschl & Osborne, 2002; NRC, 2005, 2008) and our overall focus on "how well it is working" and "what still needs to be done." Finally, we also need to acknowledge that the classroom social norms outside the laboratory setting and the number of ideas available to students changed over the course of the intervention. These factors, therefore, might have also contributed to the observed differences in ways students interacted with each other, materials, and ideas. With these three limitations in mind, we can now present our tentative answers to the four research questions that we posed at the beginning of this article.

To What Extent Does a Series of Laboratory Activities Designed Using the ADI Instructional Model Influence the Ways Students Participate in Scientific Argumentation and Craft a Written Scientific Argument?

Our findings indicate that the ADI instructional model seems to function, for the most part, as predicted by our theoretical and empirical framework. First, the results of our analysis indicate that the students' ability to participate in scientific argumentation in a manner that reflects the cognitive, epistemic, and social norms of science (i.e., disciplinary engagement) improved over the course of the intervention. Students in five of the six groups, for example, were much more likely to discuss ideas when they were introduced into the conversation and challenged the ideas of others with greater frequency after the intervention (see Figures 3 and 4). In addition to these indicators of better engagement, the students in five of the six groups also used criteria valued in science, such as fit with data or adequacy of an explanation, more frequently postintervention than they did prior to the intervention (see Figure 5). This observation suggests that the nature of the argumentation that the students engaged in at the end of the semester was more disciplinary in nature than it was at the beginning.

It is important to point out that the students in this study did not abandon using informal criteria altogether after the intervention; nor do we think that these students should have abandoned using this type of criteria as a way to evaluate ideas as a result of the intervention. The use of informal criteria, such as how plausible an idea is or how well an idea fits with personal experiences, can play an important role in scientific argumentation because these type of criteria, when coupled with an adequate level of content knowledge about the phenomenon under investigation, can serve as a useful and productive way to eliminate flawed explanations or ideas. The results of our analysis, however, does indicate that these students used the rigorous criteria that were emphasized during each ADI investigation with greater frequency after the intervention and thus seemed to privilege some of the criteria that are valued in science more than they did at the beginning of the semester.

Our analysis also indicates that all the groups were able to generate higher quality written arguments after the intervention (see Figure 7). All six groups included a more sufficient explanation postintervention and used better evidence and reasoning in their argument to support their ideas. Although the conceptual quality of the explanations the students included in the arguments did not improve much, the results of this study indicate that these students developed a better understanding of what counts as an explanation, evidence, and reasoning over the course of the intervention. It is important to note, however, that we did not assess the students' understanding of why it is important to include evidence and reasoning in a scientific argument nor did we have the groups generate arguments without using prompts to encourage them to include both evidence and reasoning in their answers pre- or postintervention. Thus, it is possible that the students simply developed a better understanding of what counts as an explanation, evidence, and reasoning in this context rather than more fluid "grasp of practice" (Ford, 2008) that will allow them to transfer their understanding of argumentation and arguments in science to other contexts. We, however, believe that developing a basic understanding of "what counts" is an important first step for students and a valuable educational outcome. After all, if students do not have a basic understanding of what counts as evidence or reasoning in a scientific argument (as was the case for these students at the beginning of the intervention), then it is highly unlikely that students will be able to provide genuine evidence or reasoning in support of their claims with or without encouragement and be able to identify invalid evidence or faulty reasoning in other contexts.

Overall, we believe that these two findings are important. They suggest that a series of laboratory activities designed using the ADI instructional model, which provides opportunities for students to participate in authentic scientific practices, encourages students to use specific criteria to evaluate the merits of ideas and provides students with educative feedback about their performance during each lab, can help some students develop new knowledge and skills. These results, especially in light of the substantial literature that indicates that students tend to struggle with many aspects of scientific argumentation (Berland & Reiser, 2009; Jimenez-Aleixandre & Erduran, 2007; Jimenez-Aleixandre et al., 2000; Osborne et al., 2004) and do not produce written arguments that reflect what counts as high quality in science (McNeill et al., 2006; Sampson & Clark, 2008; Sandoval & Millwood, 2005), suggest that this instructional model has great promise and potential. Yet, despite these promising findings, we want to stress that this study was exploratory in nature and the lack of a control group and the small sample size limits the generalizability of these findings. Nonetheless, the results reported here indicate that an efficacy study of the model with a larger sample and a control group is warranted.

Is There a Relationship Between the Ways Groups of Students Participate in Scientific Argumentation and the Nature of the Written Arguments They Create?

Our findings, as noted earlier, indicate that there seems to be a relationship between the way these students participated in scientific argumentation and the nature of the written arguments they created. Groups that had higher levels of disciplinary engagement in scientific argumentation also crafted higher quality written arguments (see Figure 8). We also did not observe any cases where a group had a high level of disciplinary engagement in scientific argumentation but produced a weak written argument or had a low level of disciplinary engagement but produced a strong argument pre- or postintervention. These observations, when taken together, indicate that there seems to be a positive correlation between these two outcome measures. However, we do not think that improved performance in one practice directly leads to a better performance in the other; instead the students seem to use the same kinds of knowledge to guide how they engage in both practices.

We suspect that students use the same kinds of knowledge to engage in scientific argumentation and to craft a scientific argument because disciplinary engagement in both practices requires a basic understanding of the epistemological commitments of science (Duschl, 2008). Students, as discussed earlier, often do not understand what counts as an argument, an explanation, evidence, reasoning, or even data in the context of science (McNeill & Krajcik, 2007; Sampson & Clark, 2008; Sandoval & Millwood, 2005). Yet, this does not mean that these terms are completely foreign to students; children simply have their own personal understanding of what these terms mean based on how they are used in other contexts (Pontecorvo & Girardet, 1993; Resnick et al., 1993; Sampson & Clark, 2009b; Stein & Miller, 1993). Students, therefore, must rely on their everyday understanding of argument and argumentation, which is based on their past experiences, when a teacher first asks them engage in these practices. Most students also do not understand (or at least privilege) the criteria or ground rules the shape how explanations or arguments are critiqued and evaluated during an episode of scientific argumentation before they learn about them in school science. As a result, most students tend to rely on the criteria they use in nonscience contexts to evaluate explanations or arguments about a natural phenomenon. Although these criteria are important and valuable in a wide range of contexts (including many science classrooms), some are not well aligned with the types of criteria that are valued in mainstream science. As a result, most students do not engage in argumentation or craft a written argument in a manner that reflects the norms and epistemological commitments of science. This was clearly the case in this study. At the beginning of the intervention, the students relied on criteria that are often used in nonscience contexts to evaluate ideas and did not include genuine evidence to their support claims in their arguments they crafted even when responding to prompts on an answer sheet.

Our current hypothesis, given the result of this analysis, is that our explicit focus on "what counts" in science led to an epistemic shift for most, but not all, of the participants in this study. As a result of this epistemic shift, the students had higher levels of disciplinary engagement in scientific argumentation and produced better arguments at the end of the intervention. We view an epistemic shift as a point in time when an individual adopts and begins to use a new framework for looking at and making sense of the world. Not unlike a paradigm shift, we see an epistemic shift as a fundamental change in the standards or criteria that an individual uses or privileges to determine what counts as warranted knowledge and how such knowledge can be generated and validated in a given context. We conjecture that an epistemic shift requires two conditions to occur. First, an individual must be introduced to new criteria or standards for what counts as warranted knowledge in an explicit fashion. Second, individuals need to be encouraged by others to use these new criteria and standards in a context where the use of these new criteria or standards are valuable and make sense. An epistemic shift, however, does not seem to be an evolutionary process given the observations we made during this study. Instead, it seems result from reaching a tipping point.

All of the students in this study, for example, were introduced to standards that can be used to determine what counts as warranted knowledge in the context of science during the first lab activity. However, few if any of these students seemed adopt these standards as their own at this point in time. Instead, the students were repeatedly exposed to and encouraged to use these new criteria to generate explanations, craft arguments, and critique each other's ideas during each laboratory activity. To facilitate this process, we used the ADI instructional model to create a classroom culture that was more conducive to student engagement in the practices of science than a more traditional laboratory setting. As a result of the sustained focus on the epistemic and social aspects of science over the entire course of the intervention, most students seemed to reach a personal tipping point, and as a result, underwent an epistemic shift. At this point in time, these students seemed to adopt the criteria privileged inside the classroom as their own and begin to use them with much greater frequency. This new epistemic framework or shared knowledge of "what counts," in turn, seemed to change the ways students interacted with each other, materials, and ideas in this context. It also seemed to change how most of the students co-constructed their arguments. This explanation for relationship between argument and argumentation, however, is only speculative at this time and will require more targeted research to substantiate.

What Types of Learning Issues Need to be Addressed to Better Help Students Learn How to Engage in Scientific Argumentation and Craft Written Scientific Arguments?

The analytical approach we used in this study enabled us to identify two learning issues that will need to be addressed to better promote and support the development of the knowledge and skills needed to participate in scientific argumentation. First, as noted earlier, it seems that the students in this study were unable or unwilling to use scientific theories, models, or laws as a tool to make sense of a natural phenomenon and to evaluate scientific knowledge. Second, groups do not always discuss a wide range of ideas and the actions of a group seem to reflect a collective confirmation bias. These two issues seem to arise when laboratory activities are designed to engage students in scientific practices (e.g., designing investigations, argumentation, and peer review) and can act as a barrier to greater student achievement. In the paragraphs that follow, we will discuss these learning issues in greater detail.

Students Do Not Use Scientific Explanations as a Tool to Make Sense of Natural Phenomena or to Evaluate Scientific Knowledge. It seems that the participants in this study did not use scientific theories, models, or laws to solve problems before or after the intervention. Our analysis, for example, indicates that the participants in this study rarely, if ever, used one of the relevant scientific theories or processes (e.g., the kinetic-molecular theory of matter or the process of combustion) or laws (e.g., the gas laws or the law of conservation of matter) introduced in class as a way to make sense of the candle and the inverted flask problem or to critique the merits of a potential explanation. These students, instead, seemed to rely more on everyday explanations (e.g., "fire needs oxygen or it goes out") rather than scientific ones (e.g., "oxygen combines with carbon during the process of combustion") or past experiences that occurred outside the classroom as a way to make sense of the phenomenon under investigation. This indicates that this instructional model, which was designed to encourage students to use scientific theories, models, and laws as a way to make sense of natural phenomenon, did not have much of an impact on this aspect of scientific argumentation.

This observation is troubling given the emphasis that was placed on this important aspect of scientific argumentation throughout the intervention. For example, students were encouraged to use theoretical criteria, such as how well a potential claim or explanation fits with other theories and laws, during the argumentation sessions and the double-blind peer review of the reports. Students were also directed to use a scientific concept or explanation introduced in class (such as molar mass or types of chemical reactions) to solve a problem (identify an unknown powder or the products of a reaction) during 5 of the 15 labs. This observation does, however, help to explain why the groups continued to generate inaccurate explanations for the candle and inverted flask problem after the intervention. For example, the idea that the flame uses up the oxygen in the flask and the loss of oxygen creates a partial vacuum was a common idea discussed by the students both pre- and postintervention. This is a reasonable inference to make based on observations alone. This idea, however, is inconsistent with the law of conservation of matter and the process of combustion. Therefore, it is not surprising that the students produced arguments with inaccurate explanations that were well supported with evidence and reasoning because the students did not take into account the theories, laws, and models of science to help them make sense of their observations or to critique the merits of their ideas.

The underlying reason for this issue, unfortunately, remains unclear and will require more research to straighten out. We can, however, suggest two potential explanations as initial candidates for exploration at this point in time. First, it is possible that these students did not understand the gas laws, combustion, the kinetic-molecular theory of matter, or the law of conservation of mass well enough to use these ideas in a novel context. This explanation, however, seems unlikely given the continual focus on these ideas throughout the semester. The second potential explanation, which we feel is more likely than the first given the content of the curriculum, is the students were not encouraged to use scientific theories, models, or laws to explain novel phenomenon enough throughout the course. As a result, students did not learn to use scientific explanations as a tool to make sense of the unknown event though they were encouraged to use theoretical criteria to evaluate explanations or claims throughout the intervention and encouraged to apply a specific concept to solve a problem during several different labs. Regardless of the underlying cause, however, the results of this study indicate these students did not use scientific theories or laws to make sense of their observations or as a way to critique the validity or acceptability of a potential explanation. This is a major issue that will need to be addressed to help students learn how to generate novel explanations and participate in argumentation in a more scientific manner.

Groups Do Not Always Discuss a Wide Range of Ideas and Their Actions Seem to Be Influenced by a Confirmation Bias. These groups of students, as noted earlier, voiced a wide range of unique content-related ideas (minimum = 9, maximum = 19) when they were engaged in scientific argumentation. Yet, many of these ideas were rejected or accepted outright instead of being discussed by the students. This type of response was common in the preintervention discussions. At the end of the intervention, however, all of the groups with the exception of Group 2 were responding to ideas by discussing them with much greater frequency (see Figure 3). Group 2 also had the lowest levels of disciplinary engagement in argumentation and crafted the weakest argument postintervention (see Figure 8). We conjecture that this is one reason why Group 2 lagged behind the other groups in terms of performance. The students in Group 2 never discussed a wide range of ideas. To illustrate this issue, consider the following excerpt taken from Group 2 after the intervention. This conversation took place immediately after the students finished reading the instructions for the task.

Student 2-B: Ok, so do you want to do the experiment first?
Student 2-C: Well, yeah. So ... [Student 2-B lights candle, places it in pan. Student 2-C puts the flask over the top of the candle. The flame goes out and water rises into the inverted flask.]
Student 2-A: That never gets old.
Student 2-C: Ok, so the flame consumes all the air because it's an enclosed area, so there's only so much air inside the bottle.
Student 2-B: So, why does the water rush up?
Student 2-C: Because there's no air so it's creating a partial vacuum. Think like a ...
Student 2-A: Or oxygen.
Student 2-C: Right. So there is no oxygen in the bottle, creating a partial vacuum, which pulls the water up.
Student 2-B: Ok. So, what is our evidence?
Student 2-C: We need some way to prove the fire does actually consume oxygen, although it's kind of evidence and proof in itself.
Student 2-A: Ok.
Student 2-B: We could try putting two candles in there, so there are two flames, and see if the water rises faster. That would prove it.

These students clearly did not discuss a wide range of ideas before agreeing on the best way to explain their observations. The performance of Group 2 also seemed to be hampered, like some of the other groups in this study, by a collective confirmation bias. A confirmation bias is a tendency to only seek out or acknowledge information that affirms an existing idea or belief (Zeidler, 1997). Many of the students in this study seemed to share the same ideas about how to explain their observations and because of a confirmation bias they neglected to explore any potential alternatives. The students in Group 2, for example, only looked for a way to support their explanation and did not try to evaluate the merits of the ideas found in their initial explanation, or for that matter, any other potential explanations. These students, in other words, were only interested in finding a way to "prove" that their explanation was correct.

This observation is once again troubling given the emphasis that was placed on the importance of discussing and testing alternative explanations throughout the intervention. Although it was clear that most students in this study understood the need to use evidence to support their explanation in the context of science (see Figure 7), it seems that some students never thought about attempting to evaluate their ideas based on the available evidence when they were asked to solve the candle and the inverted flask problem. This issue is especially problematic when everyone in a group has similar ideas and no one values the importance of testing them (as seemed to be the case in Group 2). We feel that this is a second major issue that will need to be addressed to help students develop the skills and habits of mind needed for productive participation in the practices of science.

IMPLICATIONS FOR THE TEACHING AND LEARNING OF SCIENCE

The development of the knowledge and abilities needed to participate in scientific argumentation and to craft written arguments involves much more than grouping students together and asking them to develop an evidence-based argument or explanation for a natural phenomenon. It also requires a focus on the discourse of science and an understanding of "what counts" in this context (Duschl, 2008; Sandoval & Reiser, 2004). The development of the knowledge and abilities needed to engage in scientific argumentation and to craft written arguments, therefore, is an inherently social and epistemic process as well as a conceptual and cognitive one. One way to promote and support this type of learning in the school science laboratory is to develop new instructional models that focus on scientific content, scientific processes, epistemology, and social norms at the same time. When aspects of science are brought together as complementary elements of instruction, as suggested by Duschl et al. (2007), rather than being treated as independent parts of a curriculum, learners can begin to develop powerful scientific ideas and habits of mind by generating and testing knowledge claims and using their understandings of the epistemological commitments of science to guide and evaluate those processes.

The results of this study support this notion. Our findings indicate that students, at least in this context, can learn how to participate in argumentation and co-construct written arguments in a manner that reflects the norms and goals of the scientific community after completing a series of laboratory activities that were designed to be more authentic and educative. The students in this study, for example, had better disciplinary engagement in scientific argumentation and produced higher quality scientific arguments than they did prior to the intervention. On the other hand, the results of this study also indicate that several learning issues persisted even when the ADI instructional model was used over an extended period of time. Many students in this study, for example, did not use scientific explanations as a tool to solve problems or to evaluate claims and some students seemed to be reluctant to discuss a wide range of ideas when they participated in an episode of scientific argumentation.

Our findings also contribute insights to science educators looking for ways to cultivate scientific argumentation inside the classroom and ways to improve students' knowledge and skills over time. Rather than teaching students specific discourse strategies or rules for engaging in argumentation or crafting arguments in a decontextualized and mechanistic manner prior to learning content, teachers can use instructional models, such as ADI, to provide a context for students to learn important content and how to participate in important scientific practices such as argumentation at the same time. Having students construct an explanation or argument as part of an investigation, for example, requires students to clarify their thinking, to generate examples, to recognize the need for additional information, and to monitor and repair gaps in their understanding. It also requires students to learn and use the criteria by which these explanations or arguments will be judged or evaluated. This type of approach, as we demonstrated here, can be an effective way to help students develop the abilities needed to participate in scientific argumentation, understand how to craft written arguments, and learn important content.

Teachers, however, will need to focus on more than "what we know" when instructional models such as ADI are used inside the classroom. Teachers will also need to focus on issues such as "how do we know what we know," "why do we believe what we know," and "what should we do to find out" inside the school science laboratory. Thus, a major challenge for science teachers will be to strike an appropriate balance between these different but important foci. Science teachers will also need to know much more than the theories, laws, and concepts of science to support and promote student learning in this type of context. Teachers that chose to use the ADI instructional model, or a model similar to it, will need to know how to manage the ideas and information that are generated by students. Teachers will also need to know how to establish and maintain a classroom culture and discourse environment inside the laboratory that is more aligned with how knowledge is communicated, represented, and argued in science. A challenge for science teacher educators in the years to come, therefore, will be to determine how to best prepare teachers so they are ready to teach in this manner. Although instructional models, such as ADI, can provide a useful tool for both science teachers and science teacher educators looking to reform laboratory-based instruction, this type of strategy is by no means a solution to all these issues.

In closing, our findings provide new insight for science educators and instructional designers interested in promoting and supporting argumentation inside the classroom. This study also demonstrates what is possible in the classroom when laboratory activities are designed to be more authentic and educative. Much work remains to be done, however, to evaluate the efficacy of the ADI instructional model in a wider range of contexts and at a larger scale and to identify other issues that might act as barriers to student learning. Studies like this one also do not allow one to conclude that a particular instructional model, such as ADI, is the most effective way to promote and support the development of the knowledge and skills need to participate in scientific argumentation and to craft written scientific arguments. Nevertheless, this study demonstrates that laboratory activities can be designed to be more authentic and educative, what students can learn how to do in this type of learning environment, and what challenges remain. This study, in other words, helps us understand how to cultivate student learning, some potential barriers that must be taken into account by science educators, and what teaching and learning inside the school science laboratory could look like in the years to come.

REFERENCES

American Association for the Advancement of Science. (1993). Benchmarks for science literacy. Washington, DC: Author.

Anderson, C. (2007). Perspectives on science learning. In S. K. Abell & N. Lederman (Eds.), Handbook of research in science education (pp. 3–30). Mahwah, NJ: Erlbaum.

Andriessen, J., Baker, M., & Suthers, D. (2003). Argumentation, computer support, and the educational contexts of confronting cognitions. In J. Andriessen, M. Baker, & D. Suthers (Eds.), Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments (pp. 1–25). Dordrecht, The Netherlands: Kluwer.

Baker, M. J. (1999). Argumentation and constructive interactions. In P. Coirer & J. Andriessen (Eds.), Foundations of argumentative text processing (pp. 179–202). Amsterdam: University of Amsterdam Press.

Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. Journal of the Learning Sciences, 9(4), 403–436.

Barron, B. (2003). When smart groups fail. Journal of the Learning Sciences, 12(3), 307–359.

Bell, P., & Linn, M. C. (2000). Scientific arguments as learning artifacts: Designing for learning from the web with KIE. International Journal of Science Education, 22(8), 797–818.

Berland, L., & Reiser, B. (2009). Making sense of argumentation and explanation. Science Education, 93(1), 26–55.

Birk, J., & Lawson, A. (1999). The persistence of the candle-and-cylinder misconceptions. Journal of Chemical Education, 76(7), 914–916.

Brem, S. K., & Rips, L. J. (2000). Explanation and evidence in informal argument. Cognitive Science, 24(4), 573–604.

Clark, D. (2006). Longitudinal conceptual change in students' understanding of thermal equilibrium: An examination of the process of conceptual restructuring. Cognition & Instruction, 24(4), 467–563.

Clark, D., & Sampson, V. (2006). Characteristics of students' argumentation practices when supported by personally-seeded discussions. Paper presented at the annual meeting of the National Association for Research in Science Teaching, San Francisco, CA.

Cobb, P., Confrey, J., diSessa, A. A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. Educational Researcher, 32(1), 9–13.

Driver, R., Asoko, H., Leach, J., Mortimer, E., & Scott, P. (1994). Constructing scientific knowledge in the classroom. Educational Researcher, 23, 5–12.

Driver, R., Leach, J., Millar, R., & Scott, P. (1996). Young people's images of science. Philadelphia: Open University Press.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. Science Education, 84(3), 287–313.

Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. Review of Research in Education, 32, 268–291.

Duschl, R., Schweingruber, H., & Shouse, A. (Eds.). (2007). Taking science to school: Learning and teaching science in grades K-8. Washington, DC: National Academies Press.

Duschl, R. A., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. Studies in Science Education, 38, 39–72.

Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners' classroom. Cognition and Instruction, 20(4), 399–483.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. Science Education, 88, 915–933.

Fleiss, J. L. (1981). Statistical methods for rates and proportions. New York: Wiley.

Ford, M. (2008). Disciplinary authority and accountability in scientific practice and learning. Science Education, 92(3), 404–423.

Hand, B., Norton-Meier, L., Staker, J., & Bintz, J. (2009). Negotiating science: The critical role of argument in student inquiry. Portsmouth, NH: Heinemann.

Hestenes, D. (1992). Modeling games in the Newtonian world. American Journal of Physics, 60, 440–454.

Hogan, K., & Maglienti, M. (2001). Comparing the epistemological underpinnings of students' and scientists' reasoning about conclusions. Journal of Research in Science Teaching, 38(6), 663–687.

Hunt, E., & Minstrell, J. (1994). A cognitive approach to the teaching of physics. In K. McGilly (Ed.), Classroom lessons: Integrating cognitive theory and classroom practice (pp. 51–74). Cambridge, MA: MIT Press.

Jimenez-Aleixandre, M., & Erduran, S. (2007). Argumentation in science education: An overview. In S. Erduran & M. Jimenez-Aleixandre (Eds.), Argumentation in science education: Perspectives from classroom-based research. Dordrecht, The Netherlands: Springer Academic.

Jimenez-Aleixandre, M., Rodriguez, M., & Duschl, R. A. (2000). "Doing the lesson" or "doing science": Argument in high school genetics. Science Education, 84(6), 757–792.

Klahr, D., Dunbar, K., & Fay, A. L. (1990). Designing good experiments to test bad hypotheses. In J. Shrager & P. Langley (Eds.), Computational models of scientific discovery and theory formation (pp. 355–401). San Mateo, CA: Morgan Kaufmann.

Kuhn, D. (1989). Children and adults as intuitive scientists. Psychological Review, 96(4), 674-689.

Kuhn, D. (1991). The skills of argument. Cambridge, England: Cambridge University Press.

Kuhn, L., & Reiser, B. (2005). Students constructing and defending evidence-based scientific explanations. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Dallas, TX.

Kuhn, L., & Reiser, B. (2006). Structuring activities to foster argumentative discourse. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Lawson, A. (1999). What should students learn about the nature of science and how should we teach it? Journal of College Science Teaching, 28, 401–411.

Lawson, A. (2002). Sound and faulty arguments generated by pre service biology teachers when testing hypotheses involving unobservable entities. Journal of Research in Science Teaching, 39(3), 237–252.

Linn, M. C., & Eylon, B.-S. (2006). Science education: Integrating views of learning and instruction. In P. Alexander & P. H. Winne (Eds.), Handbook of educational psychology (pp. 511–544). Mahwah, NJ: Erlbaum.

McNeill, K. L., & Krajcik, J. (2007). Middle school students' use of appropriate and inappropriate evidence in writing scientific explanations. In M. Lovett & P. Shah (Eds.), Thinking with data: The proceedings of 33rd Carnegie Symposium on Cognition. Mahwah, NJ: Erlbaum.

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. Journal of the Learning Sciences, 15(2), 153–191.

Minstrell, J. (2000). Student thinking and related assessment: Creating a facet-based learning environment. In N. Raju, J. Pellegrino, M. Bertenthal, K. Mitchell, & L. Jones (Eds.), Grading the nation's report card: Research from the evaluation of NAEP. Washington, DC: National Academy Press.

National Research Council. (1996). National Science Education Standards. Washington, DC: National Academy Press.

National Research Council. (2005). America's lab report: Investigations in high school science. Washington, DC: National Academy Press.

National Research Council. (2007). Taking science to school: Learning and teaching science in grades K-8. Washington, DC: National Academies Press.

National Research Council. (2008). Ready, set, science: Putting research to work in K-8 science classrooms. Washington, DC: National Academy Press.

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in science classrooms. Journal of Research in Science Teaching, 41(10), 994–1020.

Passmore, C., & Stewart, J. (2002). A modeling approach to teaching evolutionary biology in high schools. Journal of Research in Science Teaching, 39(3), 185–204.

Peckham, G. (1993). A new use for the candle and tumbler myth. Journal of Chemical Education, 70(12), 1008–1009.

Pontecorvo, C., & Girardet, H. (1993). Arguing and reasoning in understanding historical topics. Cognition and Instruction, 11(3 & 4), 365–395.

Reiser, B. J., Tabak, I., Sandoval, W. A., Smith, B. K., Steinmuller, F., & Leone, A. J. (2001). BGuILE: Strategic and conceptual scaffolds for scientific inquiry in biology classrooms. In S. M. Carver & D. Klahr (Eds.), Cognition and instruction: Twenty-five years of progress (pp. 263–305). Mahwah, NJ: Erlbaum.

Resnick, L. B., Salmon, M., Zeitz, C. M., Wathen, S. H., & Holowchak, M. (1993). Reasoning in conversation. Cognition and Instruction, 11(3 & 4), 347–364.

Roediger, H., & Karpicke, J. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. Psychological Science, 17(3), 249–255.

Salomon, G. (1993). Distributed cognitions: Psychological and educational considerations. New York: Cambridge University Press.

Sampson, V., & Clark, D. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. Science Education, 92(3), 447–472.

Sampson, V., & Clark, D. (2009a). A comparison of the collaborative scientific argumentation practices of two high and two low performing groups. Research in Science Education, online first. Available online at http://www.springerlink.com/content/108230/.

Sampson, V., & Clark, D. (2009b). The effect of collaboration on the outcomes of argumentation. Science Education, 93(3), 448–484.

Sampson, V., Walker, J., Dial, K., & Swanson, J. (2010, March). Learning to write in undergraduate chemistry: The impact of Argument-Driven Inquiry. Paper presented at the 2010 Annual International Conference of the National Association of Research in Science Teaching (NARST), Philadelphia, PA.

Sandoval, W. A., & Millwood, K. (2005). The quality of students' use of evidence in written scientific explanations. Cognition and Instruction, 23(1), 23–55.

Sandoval, W. A., & Reiser, B. J. (2004). Explanation driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. Science Education, 88(3), 345–372.

Saul, E. W. (Ed.). (2004). Crossing borders in literacy and science instruction: Perspectives on theory and practice. Arlington, VA: NSTA Press.

Schauble, L., Klopfer, L. E., & Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. Journal of Research in Science teaching, 28, 859–882.

Schofield, J. (1990). Increasing the generalizability of qualitative research. In E. W. Eisner & A. Peshkin (Eds.), Qualitative inquiry in education: The continuing debate (pp. 201–232). New York: Teachers College Press.

Scott, P. H., Asoko, H., & Leach, J. (2007). Students conceptions and conceptual learning in science. In S. K. Abell & N. Lederman (Eds.), Handbook of research in science education (pp. 31–56). Mahwah, NJ: Erlbaum.

Shanahan, C. (2004). Better textbooks, better readers and writers. In W. Saul (Ed.), Crossing borders in literacy and science instruction: Perspectives on theory and practice. Arlington, VA: NSTA Press.

Stein, N. L., & Miller, C. (1993). The development of memory and reasoning skill in argumentative contexts: Evaluating, explaining, and generating evidence. In R. Glaser (Ed.), Advances in instructional psychology (Vol. 4, pp. 235–285). Hillsdale, NJ: Erlbaum.

Stewart, J., Cartier, J. L., & Passmore, C. (2005). Developing understanding through model-based inquiry. In S. Donovan & J. D. Bransford (Eds.), How students learn science in the classroom. Washington, DC: The National Academies Press.

Stigler, J. W., Gonzales, P., Kawanaka, T., Knoll, S., & Serrano, A. (1999). The TIMSS videotape classroom study: Methods and findings from an exploratory research project on eighth-grade mathematics instruction in

Germany, Japan, and the United States. Washington, DC: U.S. Department of Education, National Center for Educational Statistics.

Suchocki, J. (2000). Conceptual chemistry: Understanding our world of atoms and molecules (1st ed.). Upper Saddle River, NJ: Benjamin Cummings.

Tabak, I., Smith, B. K., Sandoval, W. A., & Reiser, B. J. (1996). Combining general and domain-specific strategic support for biological inquiry. In C. Frasson, G. Gauthier, & A. Lesgold (Eds.), Proceedings of the Third International Conference on Intelligent Tutoring Systems (ITS '96). Montreal, Québec, Canada: Springer-Verlag.

Wallace, C., Hand, B., & Prain, V. (Eds.). (2004). Writing and learning in the science classroom. Boston, MA: Kluwer.

Wallace, C., Hand, B., & Yang, E.-M. (2005). The science writing heuristic: Using writing as a tool for learning in the laboratory. In W. Saul (Ed.), Crossing borders in literacy and science instruction. Arlington, VA: NSTA Press.

Weiss, I. R., Banilower, E. R., McMahon, K. C., & Smith, P. S. (2001). The 2000 national survey of science and mathematics education. Chapel Hill, NC: Horizon Research.

Wells, M., Hestenes, D., & Swackhamer, G. (1995). A modeling method for high school physics instruction. American Journal of Physics, 63(7), 606–619.

Zeidler, D. L. (1997). The central role of fallacious thinking in science education. Science Education, 81, 483–496.