The Impact of Collaboration on the Outcomes of Scientific Argumentation

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

DOUGLAS CLARK
College of Education, Arizona State University, Tempe, AZ 85287-0911, USA

Received 10 December 2007; revised 13 June 2008; accepted 27 June 2008

DOI 10.1002/sce.20306
Published online 5 November 2008 in Wiley InterScience (www.interscience.wiley.com).

ABSTRACT: This study examines three questions about the impact of collaboration during scientific argumentation. First, do groups craft better arguments than individuals? Second, to what degree do individuals adopt and internalize the arguments crafted by their group? Third, do individuals who work in groups learn more from their experiences than individuals who work on their own? To examine these questions, 168 high school chemistry students were randomly assigned, using a matched pair design to *collaborative* or *individual* argumentation conditions. Students in both treatment conditions first completed a task that required them to produce an argument articulating and justifying an explanation for a discrepant event. The students then completed mastery and transfer problems on their own. The results of this study indicate that (a) groups of students did not produce better arguments than students who worked alone, (b) a substantial proportion of the students adopted at least some elements of their group's argument, and (c) students from the collaborative condition demonstrated superior performance on the mastery and transfer problems. These observations indicate that collaboration was beneficial for individual learning but not for initial performance on the task. The study concludes with a discussion of these observations and recommendations for future research. © 2008 Wiley Periodicals, Inc. *Sci Ed* **93:**448–484, 2009

INTRODUCTION

Current research suggests that integrating scientific argumentation into the teaching and learning of science can promote scientific literacy (Driver, Newton, & Osborne, 2000;

Correspondence to: Victor Sampson; e-mail: vsampson@fsu.edu
Contract grant sponsor: National Science Foundation.
Contract grant number: 0334199.
Any opinions, findings, and conclusions expressed in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Duschl & Osborne, 2002; National Research Council [NRC], 2000). However, fostering productive scientific argumentation in classrooms has proven difficult because students often struggle with tasks that require them to propose, support, critique, and refine ideas (Jimenez-Aleixandre, Rodriguez, & Duschl, 2000; Kelly, Druker, & Chen, 1998; Kuhn & Udell, 2003; Osborne, Erduran, & Simon, 2004; Sandoval & Millwood, 2005; Zeidler, 1997). To address this challenge, many researchers (e.g., Abell, Anderson, & Chezem, 2000; Bell & Linn, 2000; Kuhn & Reiser, 2005; McNeill, Lizotte, Krajcik, & Marx, 2006; Schwarz & Glassner, 2003) have encouraged students to work in collaborative groups when they engage in scientific argumentation. The work of these authors suggests that opportunities to collaborate with others can lead to more productive scientific argumentation and improved learning outcomes because groups can pool knowledge and take advantage of different cognitive or monitoring resources. Few studies, however, have explicitly compared individual and group performance on tasks that require students to engage in scientific argumentation or examined the benefits of collaboration during scientific argumentation for individual learning.

OBJECTIVES

Given this gap in the literature, the current study examined three important questions about the impact of collaboration as a means for fostering more productive argumentation in science classrooms. First, do groups craft better arguments than individuals? Second, to what degree do individuals adopt and internalize the arguments crafted by their group? Third, do individuals who work in groups learn more from their experiences than individuals who work on their own? The overall goal of this study, therefore, was to examine the value of collaborative as a way to foster more productive scientific argumentation inside the classroom and as a way to improve learning outcomes for individual students. In the following sections, we provide an overview of the research that served as the impetus for this study and the theoretical framework that informed our work.

LITERATURE REVIEW

Scientific Argumentation in Science Education

An explicit goal of the current reform movement in science education is to promote scientific literacy in the United States (American Association for the Advancement of Science [AAAS], 1993; NRC, 1996). One way to encourage scientific literacy is to help students develop a better understanding of science subject matter in terms of the declarative knowledge specifically associated with the physical, life, and earth sciences. In addition to helping students develop this type of knowledge, however, science education programs designed to promote true scientific literacy also need to help learners understand how this knowledge is generated, justified, and evaluated by scientists and how to use such knowledge to engage in inquiry in a manner reflecting the practices of the scientific community (Driver et al., 2000; Duschl & Osborne, 2002). In other words,

> Learning science involves young 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, Asoko, Leach, Mortimer, & Scott, 1994, p. 8)

Inquiry is at the heart of current efforts to help students develop this type of scientific literacy (AAAS, 1993; NRC, 2000). In this literature, inquiry is often described as a knowledge-building process in which explanations are developed to make sense of data and then presented to a community of peers so they can be critiqued, debated, and revised (Driver et al., 2000; Duschl, 2000; Sandoval & Reiser, 2004; Vellom & Anderson, 1999). Thus, the ability to participate in productive scientific argumentation (i.e., the ability to examine and then either accept or reject the relationships or connections between and among the evidence and theoretical ideas invoked in an explanation) is viewed by many as an indicator of scientific literacy (Driver et al., 2000; Duschl & Osborne, 2002; Jimenez-Aleixandre et al., 2000; Kuhn, 1993; Siegel, 1989). Yet opportunities for students to learn how to engage in productive scientific argumentation in the context of science are rare (Newton, Driver, & Osborne, 1999; Simon, Erduran, & Osborne, 2006). It is, therefore, not surprising that empirical research examining ways to promote argumentation in science classrooms and to support students as they learn how to engage in this complex practice has intensified over the past decade.

A great deal of this research has focused on the development of new pedagogical practices (e.g., Duschl, Ellenbogan, & Erduran, 1999; Engle & Conant, 2002; Kuhn & Reiser, 2006; McNeill et al., 2006; Osborne et al., 2004) or technology-enhanced learning environments (e.g., Andriessen, Erkens, Van de Laak, Peters, & Coirier, 2003; Bell & Linn, 2000; Clark & Sampson, 2005, 2006b; Goldman, Duschl, Ellenbogen, Williams, & Tzou, 2002; Sandoval & Reiser, 2004; Toth, Suthers, & Lesgold, 2002) that integrate scientific argumentation into the teaching and learning of science. This research suggests that integrating argumentation into the teaching and learning of science requires a fundamental shift in the way science is taught. In general, curricula designed to promote scientific argumentation must emphasize *how science knows* in addition to *what science knows* and instructional strategies designed to support scientific argumentation need to provide opportunities for students to evaluate and critique the processes, contexts, and products of inquiry (Carey & Smith, 1993; Driver et al., 2000; Duschl, 2000; Duschl & Osborne, 2002; Kuhn & Reiser, 2006).

A common framework for encouraging students to engage in scientific argumentation inside the classroom has focused on providing opportunities for students to investigate and make sense of complex problems (e.g., Baker, 1999; Coleman, 1998; deVries, Lund, & Baker, 2002; Kelly & Chen, 1999; Kelly et al., 1998; Kuhn & Reiser, 2005; McNeill & Krajcik, 2007; Rochelle, 1992). One way to accomplish this task involves engaging students in the production and evaluation of *intellectual artifacts*. Intellectual artifacts can take a number of forms (such as an article, poster, concept map, mathematical equation, data table, or graph) and are used by individuals to represent and communicate valued forms of knowledge (Roth, McGinn, Woszcyna, & Boutonne, 1999). This approach frames the goal of inquiry as the construction of a good intellectual artifact that provides an underlying explanation that explicates or describes the phenomenon under investigation or an argument that provides and justifies an explanation. Students develop one or more ways to investigate the phenomenon, make sense of the data they gather, and produce an intellectual artifact that makes their understanding visible. The quality of these artifacts then becomes the focal point of discussion in the classroom as students evaluate and critique methods, explanations, evidence, and reasoning.

Another common framework for promoting and supporting scientific argumentation in classrooms has focused on designing activities or tasks that require students to examine and evaluate alternative theoretical interpretations of a particular phenomenon (Monk & Osborne, 1997; Osborne et al., 2004). This type of approach provides opportunities for students to examine competing explanations, evaluate the evidence that does or does not support each perspective, and construct arguments justifying the case for one explanation or another (e.g., Linn & Eylon, 2006; Osborne et al., 2004; White & Gunstone, 1992). Linn and Eylon (2006) and White and Gunstone (1992) suggest that this type of instructional approach not only provides opportunities for students to evaluate alternative ideas but also encourages students to use evidence to distinguish among these ideas in a more rational way. Osborne et al. (2004) suggests that this type of approach is an effective way to promote learning from scientific argumentation because it represents a "view developed in literacy studies that secure knowledge and understanding are as much a product of knowing why some ideas are erroneous as much as why other ideas are correct" (p. 997).

How Students Engage Scientific Argumentation

Argumentation, in everyday contexts, is often described as a process of debate or a discussion between people with different viewpoints. Numerous studies have shown that people have a natural ability to engage in this type of discourse. People, young, and old know how to justify, defend, and attack a viewpoint during conversation because this type of activity enables them to resolve conflict or to achieve various goals (Eisenberg & Garvey, 1981; Schwarz & Glassner, 2003; Stein & Bernas, 1999; Stein & Miller, 1991). In fact, studies of this type of argumentation show that students are skillful at supporting their ideas, challenging, and counterchallenging points during conversations that focus on everyday issues (e.g., Baker, 1999; Pontecorvo & Girardet, 1993; Resnick, Salmon, Zeitz, Wathen, & Holowchak, 1993; Stein & Miller, 1993). Others (e.g., Gallas, 1995; May, Hammer, & Roy, 2006; Sandoval, 2003b) have even described the ways students engage in argumentation as being similar to the argumentation of scientists (e.g., preferring causal explanations, being skeptical of new ideas, supporting claims with data, using analogies). Together, this literature suggests that students enter the classroom with productive cognitive resources (Hammer, 2000; Hammer & Elby, 2003) that they draw upon when asked to generate or evaluate explanations and arguments in the context of science.

Current research, however, indicates that students often struggle with scientific argumentation. This, we argue, is due to a lack of understanding of the goals and processes of scientific argumentation and how these goals and processes diverge from the forms of argumentation they are accustomed to rather than a lack of skill or natural ability. For example, students are often asked to generate an explanation for why or how something happens during activities designed to engage students in scientific argumentation. To do this, students must first make sense of the phenomenon they are studying based on the data available to them. Research suggests that this practice is often difficult for students because they often do not seek out (or generate) data that can be used to help test their ideas or discriminate between competing hypotheses (e.g., Klahr, Dunbar, & Fay, 1990; Schauble, Klopfer, & Raghavan, 1991). In addition, students often rely on their personal views rather than using the data at hand (Hogan & Maglienti, 2001). In other cases, students may use data from an investigation to draw conclusions, but not the appropriate data (McNeill & Krajcik, 2007; Sandoval & Millwood, 2005). Furthermore, students often fail to attend to important patterns in the data (Kuhn, Garcia-Mila, Zohar, & Anderson, 1995; Schauble, Glaser, Duschl, Schulze, & John, 1995; Zeidler, 1997) and are biased to ignore or distort data that threaten strongly held ideas (Chinn & Brewer, 1993; Driver et al., 1994). This body of research indicates that making sense of a phenomenon based on data, which is an important goal and aspect of scientific argumentation but not of everyday argumentation, is often challenging for students.

Current research also suggests that students have a great deal of difficulty *generating an explanation* that articulates their understanding in a way that is consistent with the types of explanations that are generated by scientists (Carey, Evans, Honda, Jay, & Unger, 1989;

Lawson, 2003; Ohlsson, 1992; Sandoval, 2003a). To construct this type of explanation, individuals must coordinate their understanding of the phenomenon under investigation with their understanding of what counts as a good scientific explanation (McNeill & Krajcik, 2007; Sandoval & Reiser, 2004; Tabak, Smith, Sandoval, & Reiser, 1996). In other words, students must understand what counts as sufficient or useful (i.e., the epistemological aspects of good explanation) and how to articulate their ideas in an appropriate manner (i.e., the rhetorical aspects of a good explanation) in science (Sandoval & Millwood, 2005). Although little is known about the criteria students use for *what counts* as a good scientific explanation, the research available indicates that students tend to offer explanations that can best be described as "observations about the world" rather than generating explanations that provide a descriptive framework or an underlying causal mechanism that can be tested empirically (Driver, Leach, Millar, & Scott, 1996). This observation, once again, seems to reflect students' lack of understanding of the goals or processes of scientific argumentation rather than their inability to articulate a viewpoint or claim.

Once they have generated a suitable explanation, students must be able to *justify their explanation* using appropriate evidence as part of the process of scientific argumentation. Research indicates that this complex task is also difficult for students (Sadler, 2004). In justifying the validity or usefulness of an explanation for a particular phenomenon, students need to be able to gather, select, and convert data into evidence to support their ideas. However, students often do not use sufficient evidence (Sandoval & Millwood, 2005) or struggle to understand what counts as evidence (Sadler, 2004). If students are confronted with large amounts of data, they often encounter difficulties differentiating between what is relevant and what is irrelevant (McNeill & Krajcik, 2007). Other research suggests that when attempting to justify an idea, many people tend to rely heavily on unsubstantiated claims to justify their perspective (Kuhn, 1991) or simply use personal inferences to replace evidence that is lacking or missing (Brem & Rips, 2000). Kuhn (1991) suggests that individuals do not use appropriate evidence to justify their explanations because most people either feel that justifying a claim with evidence is unnecessary or do not understand what counts as evidence.

This literature 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), for why they used a particular piece of evidence to justify an explanation (Bell & Linn, 2000; Erduran, Simon, & Osborne, 2004; Jimenez-Aleixandre et al., 2000). Current research suggests that even when students support their explanations with evidence, they rarely discuss the validity of such a connection (Kuhn & Reiser, 2005; McNeill & Krajcik, 2007) or make reference to the scientific principles that allow them to make that connection (Lizotte, Harris, McNeill, Marx, & Krajcik, 2003; Lizotte, McNeill, & Krajcik, 2004). Similarly, Bell and Linn (2000) found that seventh- and eighth-grade students rarely include backings to the warrants in their arguments even when they are encouraged to do so. They suggest that students "omit backings because they assume their audience already knows about them" (p. 808) and will "only provide backing to substantiate a warrant in their argument if the warrant is called into question" (p. 808). This is consistent with other research that shows that individuals often do not use scientific knowledge to support their decisions in everyday situations (Aikenhead, 2004; Linn, Eylon, & Davis, 2004; Zohar & Nemet, 2002).

Similar to a student's ability to articulate a coherent and useful explanation, providing sufficient evidence and appropriate warrants or reasoning may also be related to students' understanding of the goals and processes of scientific argumentation (Sandoval & Millwood, 2005). In other words, justifying or defending an explanation in the context of science requires an understanding of what counts as convincing and persuasive *in science* (Duschl, 1990). For example, it takes knowledge of a particular domain to recognize an explanation as unwarranted or to understand what counts as evidence in that domain. Previous research with students has found that their success at completing inquiry practices, such as using appropriate evidence and reasoning to support or refute an idea, is highly dependent on their understanding of the domain (Kenyon & Reiser, 2005; Kuhn & Reiser, 2005). In other words, if students do not understand what counts as evidence in science or are unaware of the scientific principles that are relevant to the phenomenon they are investigating, then it is highly unlikely that students will be able to apply this information when they are asked to construct an argument to propose and justify an idea.

Students also have difficulty *evaluating the validity or acceptability of an explanation* for a given phenomenon during scientific argumentation. Current research indicates that students rarely use criteria that are consistent with the standards of the scientific community to determine which ideas to accept, reject, or modify. For example, the work of Hogan and Maglienti (2001) and Linn and Eylon (2006) suggests that students often rely on inappropriate criteria such as the teacher's authority or consistency with their personal beliefs to evaluate the merits of a scientific explanation. This research suggests that students rarely use criteria such as fit with evidence, consistency with other theories, laws, or models, and predictive power. Other researchers, such as D. Kuhn (1989) and L. Kuhn and Reiser (2005), have found that students often do not differentiate between evidence and inference when attempting to warrant one explanation over another or to challenge the validity of a claim. Students, as a result, often do not base their decisions to accept or reject an idea on available evidence and appropriate reasoning. Instead students tend to use inappropriate reasoning strategies to warrant one particular view over another (Zeidler, 1997) and distort, trivialize, or ignore evidence in an effort to reaffirm their own conceptions (Clark & Sampson, 2006a; Kuhn, 1989). Therefore, as Siegel (1995) suggests, fostering more productive scientific argumentation inside the classroom may require changing the criteria students use to determine "the goodness, normative status, or epistemic forcefulness of candidate reasons for belief, judgment and action" (p. 162).

To summarize, these studies indicate that the nature of the task or what students are expected to do and how activities are structured can either facilitate or constrain opportunities for students to engage in scientific argumentation within the classroom (e.g., Andriessen, Baker, & Suthers, 2003; Duschl et al., 1999; Forman, Larreamendy-Joerns, Stein, & Brown, 1998; Kelly & Chen, 1999; Osborne et al., 2004; Sandoval & Reiser, 2004). However, in spite of the dexterity students often show at supporting and refuting viewpoints in everyday conversation, it appears that students struggle with many aspects of scientific argumentation. Furthermore, these studies show that students bring important resources for argumentation to the classroom from everyday experiences, but that students encounter difficulties making sense of data, generating appropriate explanations, justifying these explanations, and explaining their reasoning in alignment with the theoretical practices of the scientific community. When asked to evaluate alternative explanations for the same phenomenon, for example, students often do not attempt to justify one viewpoint over another and they tend to rely on inappropriate standards for what counts as quality in science when critiquing ideas. These studies, when taken together, suggest that students need time and support as they build and transition from one form of argumentation to the other.

One way to help students make this transition within the constraints of a science classroom involves encouraging students to collaborate with one another during activities or tasks that require the generation or evaluation of scientific explanations and arguments. Under these conditions, the combined creative, informational, and metacognitive resources of group members might result in a better product than what an individual could produce on his or her own. A collaborative effort might also enhance students learning *from* and *about* scientific argumentation, because students are exposed to new ideas, ways of thinking, or ways of talking or writing about the topic that they can integrate with their developing understanding of the content and the practice of scientific argumentation. Few studies, however, have attempted to empirically test these ideas about collaboration and argumentation in the context of science education. To answer this question we must therefore examine the literature that documents the impact of the collaboration on performance and learning in other contexts.

The Effects of Collaboration on Performance and Learning

Assessments of the effects of collaboration on student performance have been made across a wide variety of other tasks and domains. Rather than yielding definitive answers about the most effective way to improve student performance on academic tasks, findings are mixed (for reviews, see Cohen, 1994; Hill, 1982; Webb & Palincsar, 1996). Empirical studies have demonstrated that groups tend to perform better than individuals on tasks that require the production of electrical diagrams (Amigues, 1988), model building (Azmitia, 1988), spatial reasoning (Phelps & Damon, 1989), and mathematical problem solving (Barron, 2000; Webb, 1985). However, in studies where students were required to write stories (Goldman, Cosden, & Hine, 1992) or engage in tasks that require rote memorization (Phelps & Damon, 1989), groups do not perform any better than individuals.

These mixed results suggest that the benefits of collaboration for academic performance are task and context dependent. Although, this literature indicates that an opportunity to collaborate with others is often beneficial when the task is complex or focuses on conceptual issues (Barron, 2000; Phelps & Damon, 1989), there are few experimental studies that have specifically examined the benefits of collaboration for tasks that require students to evaluate alternative explanations and then generate an argument. To complicate matters further, current research suggests that the ways groups attempt to accomplish a task, what they discuss, and how they interact with each other when engaged in argumentation in the context of science can have a negative impact on group outcomes. For example, some groups spend a great deal of time discussing the procedural aspects of the task rather than the underlying concepts involved (Jimenez-Aleixandre et al., 2000). Individual group members can limit how other group members participate and can shape how a group attempts to accomplish an assigned task (Richmond & Striley, 1996; Southerland, Kittleson, Settlage, & Lanier, 2005). Even when groups interact in a seemingly constructive manner, individuals will often ignore ideas proposed by their classmates (Kelly & Chen, 1999; Schwarz & Glassner, 2003; Southerland, Kittleson, Settlage, & Lanier, 2005) or rely on inappropriate criteria to evaluate the quality of ideas (Cartier & Stewart, 2000; Kuhn & Reiser, 2006). Overall, this research indicates that certain interactions that take place between group members can act as barriers to productive group outcomes and that opportunities to collaborate with others may not always be valuable (Engle & Conant, 2002; Osborne et al., 2004; Richmond & Striley, 1996; Yerrick, 2000).

Another issue that remains unclear focuses on the effect of collaboration on individual learning. When students work together to solve a problem or to complete a task, it is often assumed that the benefits of collaboration result from individuals pooling knowledge, combining different ideas, integrating different cognitive strengths, and the opportunity to take advantage of improved error correction and monitoring capabilities. When engaged in scientific argumentation, for example, this means that students can share the work of making sense of counterintuitive observations, evaluating the validity or acceptability of ideas, finding and organizing data that can be used as evidence, and generating an argument that articulates and justifies a particular explanation. However, even if a group can take advantage of these resources to develop a high quality solution or product, there is no reason to expect that all of the work conducted jointly would be mastered or internalized by all individuals in the group. As Hatano and Inagaki (1991) point out, "pieces of information distributed among members can be used to solve a given problem without being coordinated into a new piece of knowledge in each member's head" (p. 335).

There are a number of studies in the literature that indicate that collaboration can lead to a better product than an individual effort, yet not result in individual learning. For example, Forman and Cazden (1985) found that dyads used more sophisticated strategies than the individuals did when they were later separated and asked to solve a similar problem. Webb (1985) also found a discrepancy between scores obtained while working in a group and scores obtained by individuals at a later time. Webb observed that students who simply copied another student's work or took a passive role received high scores during the group work but low scores on the posttest. However, students who had trouble asked for help and then worked to apply their peers' ideas learned how to solve the problems and improved as a result of the experience. Although the group outcome is all that matters in many real-world settings (e.g., Hutchins, 1995), education in school settings needs to also focus on the learning of the individuals involved (Salomon, 1993).

A number of researchers have suggested that the benefits of collaboration for individual learning do not simply arise from asking individuals to work in groups; rather it is the result of individuals engaging in certain types of learning processes during these types of activities (Cohen, 1994). These include opportunities to encounter new ideas or perspectives (Linn & Eylon, 2006; Webb & Palincsar, 1996), to resolve differing perspectives through discussion (Amigues, 1988; Phelps & Damon, 1989), to explain one's thinking about a phenomenon (King, 1990; Webb, Troper, & Fall, 1995), to provide or receive critiques (Linn & Eylon, 2006; Webb & Palincsar, 1996), to observe the strategies of others (Azmitia, 1988), and to listen to the explanations of others (Coleman, 1998; Hatano & Inagaki, 1991; Webb, 1985). This literature suggests that engaging in argumentation with others may be more beneficial for individual learning than engaging in argumentation alone because this type of activity is likely to trigger many of the learning processes highlighted above. For example, when evaluating the acceptability of alternative explanations for a given phenomenon students must explain their own thinking, listen to the explanations of others, and resolve differing perspectives through discussion. However, as previously mentioned, recent research suggests that it is often difficult to engage groups of students in productive scientific argumentation. As a result, many of the learning processes outlined above, such as resolving differences through discussion and providing critiques, may not occur when students engage in scientific argumentation with others. Moreover, many of these processes are not unique to collaboration; that is, many of these processes can also be observed when students work alone. For example, conflict and controversy can arise within oneself when one thinks more deeply about, or attempts to integrate new information with, one's prior beliefs. The resolution of such conflict can produce learning gains (Chi, 2000). Receiving explanations and providing critiques have also been shown to promote learning (Chi, Bassok, Lewis, Reimann, & Glaser, 1989), and these processes do not require individuals to interact with other people. In the end, it seems that a number of questions about the value of collaboration during scientific argumentation for individual learning remain unresolved and thus warrant further investigation.

THEORETICAL FRAMEWORK

To guide our work, we have adopted a view of argumentation in science as a knowledge building and validating practice in which individuals propose, support, critique, and refine ideas in an effort to make sense of the natural world (e.g., Driver et al., 2000; Kuhn, 1993). This perspective describes argumentation in science as a practice that is used "to solve problems and advance knowledge" (Duschl & Osborne, 2002, p. 41) rather than as an effort to "justify or refute a particular standpoint" (van Eemeren, Grootendorst, & Henkemans, 2002, p. 38), or as the articulation of informal reasoning (e.g., Perkins, Farady, & Bushy, 1991; Sadler, 2004; Zohar & Nemet, 2002). This perspective also differentiates between terms such as explanation, argument, and argumentation. Explanations are statements that explicate or describe natural phenomenon, arguments provide and justify an explanation, and argumentation is the process of generating explanations, constructing arguments, and critiquing the processes, contexts, and products of inquiry (i.e., explanations or arguments). When conceptualized in this manner, argumentation "can be seen to take place as an *individual* activity, through thinking and writing, or as a *social* activity taking place within a group" (Driver et al., 2000, p. 291 emphasis in original).

To help the participants in this study understand what counts as a high-quality argument in science and to assess the quality of the arguments generated by these students, we used a Toulmin-inspired framework similar to frameworks adopted by a number of other researchers in science education (e.g., Kuhn & Reiser, 2005; Lizotte et al., 2004; McNeill & Krajcik, 2007; Osborne et al. 2004). This framework (see Figure 1) conceptualizes a scientific argument as three interrelated components: an explanation (similar to Toulmin's claim), evidence (similar to Toulmin's data), and reasoning (a combination of Toulmin's warrants

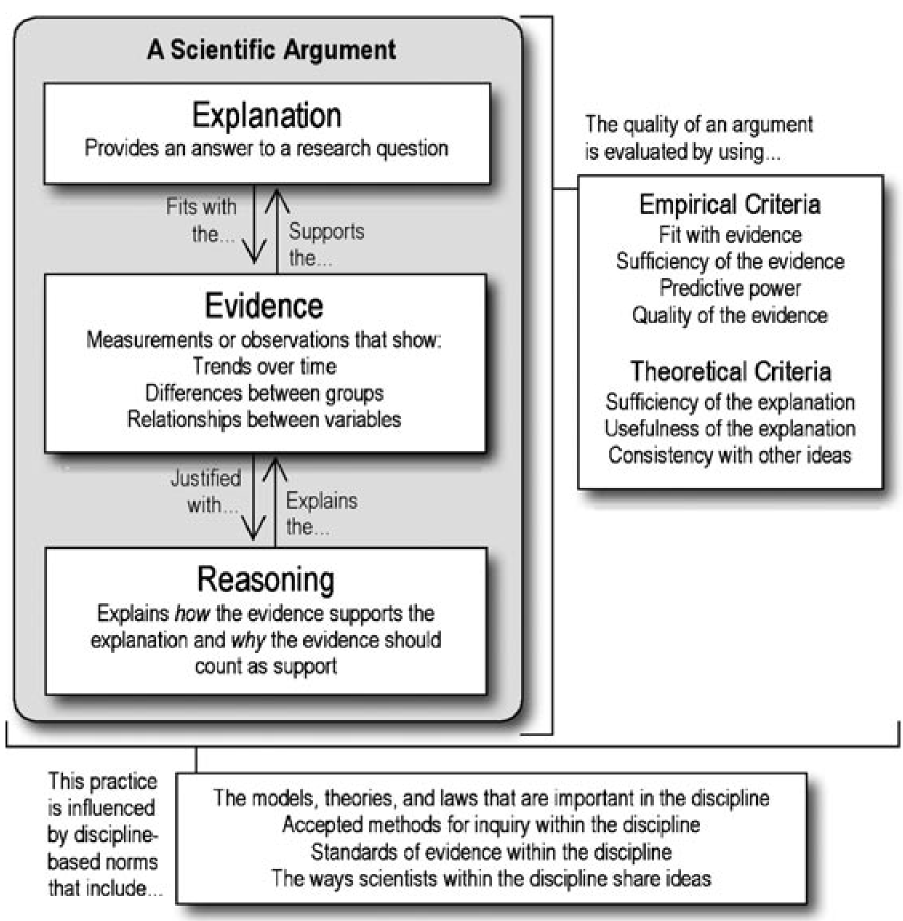


Figure 1. The argumentation framework used in this study.

and backings). The *explanation* component of the framework requires students to construct an answer to the research question that guides the investigation. Depending on the question guiding the students' investigation, this explanation can offer a solution to a problem (e.g., the unknown powder is sodium chloride), articulate a descriptive relationship (e.g., as the temperature of a gas increases, so does the volume), or provide a causal mechanism. The *evidence* component of the framework requires students to include measurements or observations to support the validity or the legitimacy of the explanation. This evidence can take a number of forms ranging from traditional numerical data (e.g., temperature of various objects in the room) to observations (e.g., the metal feels colder than wood). However, for this information to be considered as evidence it must be used to either show (a) a trend over time, (b) a difference between groups, or (c) a relationship between variables. The *reasoning* component of the framework indicates that an argument needs to include a rationalization that shows why the evidence supports the claim and why the evidence provided should count as evidence. Kuhn and Reiser (2005), Lizotte et al. (2004), and McNeill and Krajcik (2007) have adapted Toulmin's model in a similar manner to guide their curriculum development efforts. Their work (which inspired this framework) suggests that this type of approach is an appropriate and productive way to introduce students to the complex practice of generating arguments and to assess the quality of the arguments they construct.

In addition to making these structural components of a scientific argument explicit, our framework also provides students with specific criteria to evaluate the quality of an argument in science. Based on the work of Stewart and colleagues (e.g., Cartier & Stewart, 2000; Passmore & Stewart, 2002; Stewart, Cartier, & Passmore, 2005), this framework encourages students to assess the quality of an argument, using empirical criteria such as (a) how well the explanation fits with available data, (b) the predictive power of the explanation, (c) the sufficiency of the evidence cited, (d) the quality of evidence included, and theoretical criteria such as (a) how consistent the explanation is with other scientific knowledge and (b) whether the explanation is a useful way to think about the phenomenon. This component of the framework is designed to help students employ appropriate criteria for determining what counts as a quality argument in science (Hogan, 2000; Kuhn & Reiser, 2006) and to give them tools that they can use to evaluate scientific ideas in a more productive manner (e.g., Schwarz & Glassner, 2003).

RESEARCH QUESTIONS

The literature reviewed here indicates that although students face challenges engaging in scientific argumentation as documented in the literature, they also bring sufficient cognitive and monitoring resources to the classroom to benefit from these activities in terms of their learning. As a result, many researchers have suggested that students should work in collaborative groups when they engage in scientific argumentation so they can take advantage of the resources. Yet, few studies have examined the impact of collaboration on the outcomes of scientific argumentation or for individual learning. This study was therefore designed to address the following research questions that stem from the overall goal of this investigation:

1. Do groups of students craft better scientific arguments than individuals?
2. To what degree do group members adopt and internalize the argument crafted by their group?
3. Do students who engage in argumentation with others demonstrate superior performance on individual mastery and transfer tasks when compared to students who engage in argumentation alone?

METHOD

Overview of the Research Design

To assess the effects of collaboration on argumentation outcomes and individual learning, the participants in this study were asked to complete a complex task that required them to engage in argumentation to make sense of and then explain a discrepant event. This task, which is called the *ice-melting blocks* problem,required students to first determine which explanation, of six plausible alternatives or one of their own design, was the most valid or acceptable way to explain their observations using available data (i.e., knowledge building scientific argumentation). Once the participants had determined which explanation best explained the phenomenon, they then had to create a written argument that articulated and justified that explanation with appropriate evidence and reasoning (i.e., knowledge validating scientific argumentation). The participants in this study, who were enrolled in introductory chemistry at a large public high school, were randomly assigned to one of two conditions to complete this task. Students assigned to an individual argumentation condition completed this task alone. Students assigned to a group argumentation condition worked as part of a group of three (triads).

Collaboration, as noted earlier, is often viewed as a way to improve scientific argumentation outcomes, because individuals can build off each other's ideas and take advantage of different cognitive and monitoring resources. Therefore, a collaborative effort should result in an argument that is more accurate and compelling, on average, than an argument produced through an individual effort. Given this hypothesis and prediction, the effect of collaboration on argumentation outcomes was assessed by comparing the quality of the arguments produced by the triads and the individuals. Argument quality was assessed using four criteria: (a) the sufficiency of the explanation, (b) the conceptual quality of the explanation, (c) the quality of evidence, and (d) the adequacy of the reasoning (see the section dependent measures and scoring for full explanation of each measure).

The effect of collaboration on individual learning was assessed using two follow-up tasks. To assess the participant's understanding of the phenomenon in question, all of the participants were asked to complete a mastery version of the ice-melting blocks problem for a second time. For this administration of the problem, each student, regardless of the treatment condition, was required to generate his or her own written argument for the ice-melting blocks problem without a list of possible explanations to choose among. To assess the participants' ability to transfer what they had learned to a different context, each student completed a conceptually similar transfer task, the *why do objects feel different* problem. The discrepant event in this problem has the same underlying cause as the ice-melting blocks problem. As with the ice-melting blocks mastery problem, the why do objects feel different problem requires each student to individually produce a written argument that articulates and justifies an explanation for the event in question. It was predicted that students who worked as part of a triad on the ice-melting blocks problem would benefit from being exposed to new ideas, ways of thinking, and discourse practices and be able to use this knowledge when working alone, which would lead to higher average scores on the mastery and transfer problems.

Content Area

The ice-melting blocks and why do objects feel different problems focus on energy transfer, thermal equilibrium, and thermal conductivity. The ice-melting blocks core and mastery problems required the students to generate a scientific argument that explains why a piece of ice that is sitting on a block made of aluminum melts faster than a piece of ice sitting on a block made of plastic. The why do objects feel different problem required students to generate an argument that explains why objects sitting in the same room, such as a metal chair leg and a table top made of wood, feel as though they are different temperatures. The full versions of the ice- melting blocks and why do objects feel different problems are included in Appendix A.

Both of these problems have the same basic explanation: Materials with different thermal conductivities transfer thermal energy at different rates. In the ice-melting blocks problem, the ice melts faster on the aluminum block because aluminum is a better conductor of thermal energy. As a result, thermal energy transfers from the aluminum block into the ice at a higher net rate than it does from the plastic block. This is also why the two blocks feel like they are at different temperatures. Thermal energy will transfer from a person's hand into the aluminum block at a higher net rate than it will into the plastic block. People often attribute differences in thermal sensation solely to differences in temperature rather than to a combination of differences in temperature and thermal conductivity (Clark, 2006). As a result, people will often say that the aluminum block is colder than the plastic block even though they are at the same temperature. This same explanation is also used to answer the why do objects feel different problem. Objects (such as a metal chair leg) that are good thermal conductors will feel colder than objects that are poor thermal conductors (such as a wood table top) that are sitting in the same room because they transfer thermal energy at different net rates.

Participants

One hundred and sixty-eight students participated in this study. These students were all enrolled in introductory chemistry at a large suburban public high school located in the Southwest United States. Forty percent of the participants were male, and 60% were female. Approximately 71% of the students were European American, 18% were Latino/a, 4% were African American, and 7% were from other ethnic backgrounds. Ten percent of these students indicated that they spoke a language other than English at home. The students ranged in age from 15 to 17 years ($M = 15.77$, $SD = 0.62$). Forty-two percent of these students were in 10th grade, 54% were in 11th grade, and 4% were in 12th grade. These participants were drawn from six different chemistry classes. Students in classes A ($n = 30$) and B ($n = 30$) were taught by one teacher, and students in classrooms C ($n = 30$), D ($n = 30$), E ($n = 24$), and F ($n = 24$) were taught by the other teacher. These teachers described these students as being "college bound" and "highly motivated."

All the chemistry classes at this school follow the same curriculum, and the teachers often use the same instructional materials. At the time of the intervention, the teachers had finished two of the eight units outlined in the district curriculum. The first unit focused on introductory chemistry topics such as the relevancy of chemistry to everyday life, the "scientific method," measurement, significant figures, mass, density, and forms of energy. The second unit focused on the foundations of chemistry and introduced students to topics such as the molecular–kinetic theory of matter, the difference between heat and temperature, and physical and chemical changes in matter. The third unit, which they had not yet completed, introduced students to atomic structure and the periodic table. The typical instructional approach used in these classrooms was direct instruction followed by confirmatory laboratory work. Although these students had a great deal of experience working in groups during the confirmatory laboratory activities, these students did not engage in collaborative problem solving, open or guided inquiry, and scientific argumentation on a regular basis.

The decision to use a sample of students that did not engage in scientific argumentation, collaborative problem solving, or inquiry on a regular basis in this study was purposeful. This decision was made in large part because current research that examines instruction in science classrooms indicates that opportunities for students to engage in these types of activities are rare (Newton et al., 1999; Roth et al., 2006). Therefore, we felt that it was important (and more useful for science educators interested in integrating scientific argumentation into the teaching and learning of science) to examine a sample of science students that would be more representative of typical classrooms rather than a group of students that were well versed in collaboration and scientific argumentation. We also felt that it was important to work with a sample of students that was unfamiliar with scientific argumentation to highlight the various resources that students can draw upon when asked to evaluate alternative explanations and to generate a convincing or persuasive scientific argument. Thus, the results of this study can be viewed as a foundation for what students can be expected to do and learn when first asked to engage in classroom practices that are foreign to them. We plan to duplicate this study with a sample of students that engage in collaboration and argumentation on a regular basis at a later time to further evaluate the impact of collaboration on argumentation outcomes and individual learning.

Assignment to Conditions

A matched-pair random assignment procedure was used to assign students within each class to one of the two treatment conditions based on their preintervention understanding of the thermal equilibrium, thermal conductivity, and the differences between heat and temperature. During the first session of the intervention (see the section procedure) all 168 of the students who agreed to participate in the study completed the Thermodynamics Content Knowledge Test. Students who had similar scores within each class period were paired so that the students within each matched-pair had more similar scores on the Thermodynamics Content Knowledge Test than any two students from different pairs. The two students within each matched-pair were then randomly assigned (by a flip of a coin) to one of the two treatment conditions: the individual argumentation condition ($n = 84$) and the group argumentation condition ($n = 84$). A paired-samples t test was then conducted to evaluate the equivalence of student background knowledge levels *across* classes. The results indicated that there were no significant differences between the participants assigned to the group argumentation condition ($M = 14.4$, $SD = 3.4$) and the participants assigned to the individual argument condition ($M = 14.2$, $SD = 3.0$, $t(166) = 0.36$, $p = .72$). Thermodynamics Content Knowledge Test scores ranged from a low of 6 to a high of 25.

Triad Composition

Empirical research has demonstrated that the gender composition of a team can result in differential patterns of participation during a collaborative task (Webb, 1984). The triads were therefore composed of same gender students in order to simplify the interpretations of results. There were 13 male triads and 15 female triads. Students within each class were randomly assigned to a triad using a lottery system.

Procedure

Students in both treatment conditions participated in four sessions. Each session was approximately 50 minutes long, and the sessions took place on consecutive school days. The scope of each session is outlined in the paragraphs that follow.

All of the participants completed the Thermodynamics Content Knowledge Test (see Appendix B for sample items) during the first session. The Thermodynamics Content Knowledge Test (coefficient $\alpha = .81$) is designed to measure students understanding of thermal equilibrium, thermal conductivity, heat transfer, and the difference between heat and temperature. It consists of 19 items and uses a two-tiered multiple-choice format similar to those employed by Treagust (1988), Odom and Barrow (1995), and Settlage and Odom (1995). This test is a refined, extended, and expanded version of the Heat Curriculum Subject Matter Test originally developed for the Computer as Learning Partner project (see Linn & Hsi, 2000, for an overview). Over the last 10 years, these items (and variations of them) have proven valid and reliable for assessing students' understanding of concepts related to thermodynamics in a number of contexts (see Clark, 2006; Clark & Linn, 2003; Lewis, 1996; Lewis & Linn, 1994; Linn & Hsi, 2000, for example). All of the participants were able to complete this instrument in the allotted time.

The students were introduced to the argument framework during the second session. The overall goals of this session included (a) providing a target structure by highlighting the different aspects of an argument that should be included in an answer (e.g., an argument needs to include an explanation, evidence, and reasoning), (b) making explicit the criteria for what counts as quality for each aspect of an argument (e.g., quality evidence consists of observations or measurements that show a difference between groups or objects), and (c) giving the students an opportunity to develop a basic understanding of some of the criteria that are used by scientists to evaluate arguments (e.g., explanations should fit with the available data). To accomplish this task, the second session's curriculum focused on (a) defining each component of the framework, (b) making the rationale behind each component explicit, (c) modeling how to construct an argument, and then (d) providing examples of both strong and weak arguments for the students to critique.

The students completed the core ice-melting blocks problem during the third session. Students assigned to the individual argumentation condition completed this task on their own, whereas individuals assigned to the group argumentation condition completed this task in a group of three (triads). The triads worked at a table in front of a video camera so that the interactions that took place between the students as they worked could be recorded. To ensure quality video and audio recordings, the triads were moved to empty rooms during this session and an adult was present to operate the video camera. The participants assigned to the individual argumentation condition remained in their classroom and were supervised by two adults as they worked.

All of the participants received three handouts. The first handout was an answer sheet that included an introduction to the problem, the research question, and two prompting questions designed to motivate students to attend to each structural element of the argument framework (explanation, evidence, and reasoning). The second handout included six different plausible explanations for the students to examine. The third handout included information about the blocks and the molecular-kinetic theory of matter that could be used to generate the evidence necessary to justify or critique the explanations. The participants were given approximately 40 minutes to complete this task. Once the students had completed their work, they were directed to highlight the evidence they included in their argument using a highlighter.

Students were required to highlight their evidence in their argument because a number of researchers have reported that scoring written arguments produced by students using Toulmin-inspired frameworks is often difficult (Duschl, 2007; Erduran et al., 2004; Kelly et al., 1998). The main difficulty reported in the literature is that it is often hard to distinguish between data and warrants (or evidence and reasoning in this framework) in a reliable manner. To complicate matters further, students often use inappropriate information (from a scientific perspective) as evidence to support their ideas (Brem & Rips, 2000; Kuhn & Reiser, 2005; McNeill et al., 2006; Sandoval & Millwood, 2005). Therefore, to use this type of framework to score the quality of the arguments produced by students in a valid and reliable way, this methodological challenge must be resolved. One way to do this is to ask students to identify the various components of their argument prior to scoring.

Students in both treatment conditions completed the mastery version of the ice-melting blocks problem during the fourth session. Unlike the previous administration of the core version of the problem, individuals in both treatment conditions were required to complete the task on their own. This mastery administration of this problem required the students to generate their own explanation without using the list of alternative explanations and then support their explanation with evidence and appropriate reasoning. Students were given the same answer sheet that included an introduction to the problem, the research question, and the two prompting questions, and the same handout that included the information about the blocks and the molecular-kinetic theory of matter. Students were given 20 minutes to complete this task. All of the participants completed this task in the allotted time.

The participants were also asked to solve the why do objects feel different transfer problem (see Appendix A) on their own during this session. This task, which is conceptually similar to the ice-melting blocks problem, required students to generate an explanation and then support their explanation with appropriate evidence and reasoning. Students received a handout that included directions, the same prompting questions, and a data set. The students were given 20 minutes to complete this task. Once again, all of the participants were able to complete this task in the allotted time.

Dependent Measures and Scoring

Four aspects of the arguments produced by the students were scored. These were (a) the sufficiency of the explanation, (b) the conceptual quality of the explanation, (c) the quality of the evidence, and (d) the adequacy of the reasoning. Each aspect was given a score based on the presence or absence of specific components. However, rather than simply "coding and counting" (Suthers, 2006), these components were used to score the overall quality of an aspect on a four-point (0–3) scale. To increase the reliability of a decision, dichotomous keys (see Appendix C for examples) were used to reduce the subjective nature of this holistic approach. 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. To assess the reliability of this coding scheme, two coders independently scored 20% of the arguments produced by the students in this study ($n = 90$). Estimates of reliability for each aspect were then calculated using Cohen's κ. A Cohen's κ value of 0.70 or greater was considered as a satisfactory indication of interrater reliability (Bakeman & Gottman, 1986; Fleiss, 1981). Any disagreements were resolved through discussion.

We assessed the sufficiency of the explanation in the arguments by evaluating how well the explanation answered the research question. A sufficient explanation, given the problems posed in this study, needed to provide (a) a causal mechanism, (b) a description of how the objects in question are different, and (c) an account of how the causal mechanism influences these objects in terms of these differences. 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 original or "seed" explanations that were provided to the students during the intervention ranged in sufficiency quality from a low of 1 to a high of 3. Students were told that they could use one of the seed explanations, a revised seed explanation, or a completely original explanation in their argument. Fourteen percent of the arguments scored for the core version of the ice-melting blocks problem included a modified version of a seed explanation and 9% contained a completely original explanation. The reliability of this scoring scheme between the two coders, as measured by Cohen's κ, was 0.87.

The conceptual quality of the explanation was assessed using a facet analysis approach (Hunt & Minstrell, 1994; Minstrell, 2000). Clark (2006) describes facets as ideas that often lack the structure of a full explanation and can consist of nominal and committed facts, intuitive conceptions, narratives based on experiences, p-prims, or mental models at various stages of development and sophistication. These ideas in an explanation are identified and coded as accurate (e.g., objects in the same room become the same temperature, energy transfers from hot to cold objects), incomplete (e.g., objects in the same room will become a similar temperature, stating that plastics change temperature slower than metal without making a connection to conductivity), or inaccurate (e.g., cold energy travels from cool objects to warm objects, conductors attract heat, metal is a good insulator). Explanations that contained more normative ideas received higher scores than explanations composed of inaccurate ideas or explanations that contained a mixture of accurate, incomplete, and inaccurate ideas. The seed explanations provided to the students during the intervention ranged in conceptual quality from a low of 0 to a high of 3. The reliability of this scoring scheme between the two coders, as measured by Cohen's κ, was 0.81.

To assess the quality of the evidence, we examined whether appropriate and relevant evidence was used to support the given explanation. Appropriate evidence was defined as measurements (e.g., the temperature of block A is 23°C) or observations (e.g., the ice melts faster on block A) 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) 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. The reliability of this scoring scheme between the two coders, as measured by Cohen's κ, was 0.74.

We assessed the adequacy of the reasoning by determining how well the individual (or a group of individuals) linked the evidence to the explanation and justified the choice of evidence. Adequate 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. The reliability of this scoring scheme between the two coders, as measured by Cohen's κ, was 0.72.

RESULTS AND DISCUSSION

The presentation of results is divided into three subsections by research question. Each subsection includes a brief overview of the analysis, the results of the analysis, and a discussion of the findings.

Do the Triads Craft Better Scientific Arguments Than the Individuals?

To compare group and individual performance, researchers typically compare an equal number of groups and individuals (e.g., 20 three-person groups with 20 individuals). In this type of design, the average performance of the groups is compared to the average performance of the individuals. However, Laughlin, Hatch, Silver, and Boh (2006) suggest that a more informative and stringent test of group versus individual performance can be conducted by comparing n groups of size s with an equivalent number of $n \times s$ individuals (e.g., 20 groups of 3 students with 60 individuals). In this type of design, the average performance of the groups can then be compared to the average performance of the best, second best, and so forth through the sth best of an equivalent number of individuals. This enables researchers to determine whether groups are able to perform as well or better than the highest performing individuals in a study *and* to compare the performance of individuals who worked in a group with an equal number of individuals who worked alone on a follow-up task.

To generate the three comparison groups needed to evaluate the performance of the triads, individual scores on the ice-melting blocks problem were calculated and then rank ordered within each classroom. The highest scoring individuals within each class (the top 5 individuals in classrooms with 30 students and the top 4 individuals in classrooms with 24 students) were grouped together to form the top third group of individuals ($n = 28$). The individuals with the next highest scores and the individuals with the lowest scores in each class were then grouped together to form the middle third group of individuals ($n = 28$) and the bottom third group of individuals ($n = 28$), respectively. A 3 × 2 ANOVA was then conducted to evaluate the differences in group achievement levels and the effect of gender *across* classes, with group (top third, middle third, and bottom third) and gender as independent variables.The ANOVA indicated a significant main effect for the group, $F(2, 78) = 87.05$, $p < .001$, $\eta^2 = .69$, but no significant main effect for gender, $F(1, 78) = 1.9$, $p = .17$, or an interaction effect between group and gender, $F(2, 78) = 0.07$, $p = .94$. Tukey's post hoc comparisons of the group main effect indicated that all three pairwise differences were significant ($p < .01$).

The performance of the 28 triads was then compared with the performance of these three categories of individuals. A 4 × 2 ANOVA was conducted to evaluate the differences in group achievement levels and the effect of gender, with group (triads, top third individuals, middle third individuals, and bottom third individuals) and gender as the independent variables.The ANOVA indicated a significant main effect for group, $F(3, 104) = 46.03$, $p < .001$, $\eta^2 = .57$, but no significant main effect for gender, $F(1, 104) = 2.32$, $p = .13$, or an interaction effect between group and gender, $F(2, 104) = 0.04$, $p = .99$. The means and standard deviation for performance on the core ice-melting blocks problem as a function of these two factors are presented in Table 1.

Follow-up analyses to the main effect for group were then used to determine whether the triads produced better arguments than the top third, middle third, and bottom third groups of individuals. The Tukey HSD procedure was used to control for Type I error across these three pairwise comparisons. The results of this analysis indicated that the triads ($M = 7.21$, $SD = 1.64$) produced significantly better arguments than the bottom third group of individuals ($M = 4.75$, $SD = 1.14$, $p < .001$, $d = 1.74$) and arguments that were equivalent in quality to those produced by the middle third individual group ($M = 6.75$, $SD = 1.01$, $p = .54$). However, the top third group of individuals ($M = 9.39$, $SD = 1.29$) produced arguments that were significantly better than the arguments produced by the triads ($p < .001$, $d = 1.48$). Overall, these results indicate that the triads produced arguments that were comparable in quality (in terms of content and structure) to the arguments produced by

TABLE 1
The Means and Standard Deviation for Performance on the Core Ice-Melting Blocks Problem as a Function of Group and Gender

		Argument Score	
Group	Gender	*M*	*SD*
Triads	Male	7.46	1.20
	Female	7.00	1.96
	Overall	7.21	1.64
Top third individuals	Male	9.60	1.27
	Female	9.28	1.32
	Overall	9.39	1.29
Middle third individuals	Male	6.92	0.79
	Female	6.62	1.15
	Overall	6.75	1.01
Bottom third Individuals	Male	5.17	1.47
	Female	4.64	1.05
	Overall	4.73	1.14

the second highest performing individuals in the study but inferior in quality to arguments produced by the highest performing individuals (see Figure 2).

Given the substantial literature that indicates that triads should outperform individuals on a complex task like the core ice-melting blocks problem (e.g., Andriessen, Erkens, et al., 2003; Mason, 1998; Rochelle, 1992; Scardamalia & Bereiter, 1994), we conducted a second analysis to confirm these results. In this analysis, we decided to compare the performance of the 28 triads to performance of the 84 individuals directly using a between-subjects design, with a condition (triads vs. individual) and gender as independent variables. The results of this analysis indicated no significant main effect for the condition, $F(1, 108) = 0.08$, $p < .78$, gender, $F(1, 108) = 1.86$, $p = .18$, or an interaction effect between condition and

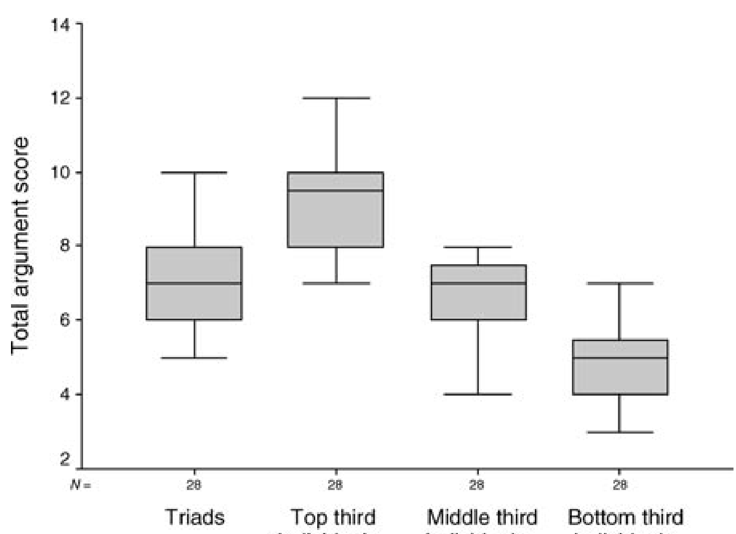


Figure 2. Performance of the triads, top third individuals, middle third individuals, and the bottom third individuals on the ice-melting blocks problem.

gender, $F(1, 108) = 0.14$, $p = .71$. The triads ($M = 7.21$, $SD = 1.64$) produced arguments that were equivalent in quality, on average, to the arguments produced by the individuals who worked alone ($M = 7.09$, $SD = 2.22$). This comparison confirms the results of the previous analysis; it seems that an opportunity to collaborate with others had little or no impact on the quality of the arguments produced by these students in response to the core ice-melting blocks problem.

This finding, however, must be interpreted in light of the intervention and the sample of students that participated in this study. As was noted in the review of the literature, the ways that group interactions are structured, students' prior knowledge, and the types of problems students are asked to solve can influence group outcomes (Alexopoulou & Driver, 1996; Barron, 2000; Phelps & Damon, 1989; Richmond & Striley, 1996). In this study, the students, who had a great deal of experience working in small groups during confirmatory laboratories but little experience with genuine collaboration and argumentation, were given a complex but well structured task. This task required them to evaluate six alternative explanations for a discrepant event and then to generate a convincing and persuasive argument that articulates and justifies the most valid or acceptable explanation. Therefore, although the collaborative effort in the current study did not result in arguments that were more accurate and compelling on average than the arguments produced by individuals, the same activity with students who had more experience engaging in collaborative scientific argumentation might demonstrate larger benefits from the activity based on their expertise and familiarity with collaboration and the goals, processes, and affordances of scientific argumentation. Similarly, this study does not discount the possibility of a more significant impact during a different type of task designed to promote or support scientific argumentation (e.g., generating an original explanation or evaluating the processes of inquiry). This study should therefore be considered as a baseline for students with minimal experience engaging in a fairly short intervention rather than as an investigation of optimized potential affordances.

We suspect that collaboration did not improve the groups' argument quality because the interactions that took place within some groups did not harness the potential collaborative opportunities for the knowledge generation and validation aspects of argumentation. This assertion is based largely on a detailed analysis of the interactions that took place between group members, ideas, and materials within two more successful and two less successful groups (Sampson, 2007; Sampson & Clark, 2008). The results of that analysis indicate that the less successful groups discussed fewer content-related ideas, were more likely to accept an idea without critical discussion when it was introduced into the conversation, relied on less rigorous criteria to evaluate the merits of an idea, and did not use the available data until they needed to generate their final argument. This suggests that some groups did not take advantage of the additional cognitive and monitoring resources available to them. If this is indeed the case, science educators interested in improving the quality of the arguments students produce will need to do more than design a meaningful task, provide students with explicit criteria for what counts as a good argument in science, and then group students together as they work. Science educators will also need to encourage students to introduce more ideas into a discussion and to value intellectual rigor, constructive criticism, and the challenging of ideas so that the interactions within a group promote better group outcomes rather than constrain them.

To What Degree Do Individual Group Members Adopt and Internalize the Arguments Produced by Their Group?

To determine the degree of adoption and internalization, the content of the arguments produced by each triad was compared to the content of the arguments that were produced by the individual group members at a later time. This was accomplished by classifying the content of each student's explanation and evidence on the second administration of the ice-melting blocks problem as either (a) the same, (b) similar, or (c) different to the explanation and evidence used by his or her triad. Explanations that were characterized as the same were defined as any response that contained all of the same components as the original explanation (e.g., using the same causal mechanism to explain why the ice melts and pointing out the same differences in the blocks). Evidence that was characterized as the same was defined as any response that used all the same pieces of information as the original evidence. A similar explanation or evidence was defined as a response that had most, but not all, of the components of the original response (e.g., leaving out one component of the explanation or substituting one piece of information highlighted as evidence used during the previous administration for another). Finally, a response that contained a majority of components that were different from the ones used in the original argument was classified as a different explanation or evidence (e.g., using a different causal mechanism and leaving out another component of the explanation or changing multiple pieces of evidence). Two coders independently scored 20% of the arguments ($n = 34$). Interrater reliability for the explanations was 82% (Cohen's $\kappa = 0.73$), and the interrater reliability for the evidence was 88% (Cohen's $\kappa = 0.82$).

This analysis is based on the premise that individual group members should generate the same argument crafted by his or her group at a later time if he or she adopted and internalized the group argument. Therefore, if the individual group members were in fact adopting and internalizing the content of the arguments produced by their group, then a relatively high proportion of these students should fall into the same or similar content categories and a relatively low proportion of these students should fall into the different content category. To determine what counts as "relatively high" and "relatively low," the proportion of students in the individual condition whose explanations and evidence on the second administration of the ice-melting blocks problem were the same, similar, and different were also categorized in the same manner. The proportion of students who fell into each category in the individual condition was then compared with the proportion of students who fell into each category in the group condition. These data are provided in Table 2.

As illustrated in Table 2, 54% of the individuals who worked individually to generate an argument for the core ice-melting blocks problem produced an argument during the second administration that contained the same basic explanation that they expressed on the first administration. Thirty-two percent produced an argument that contained a similar explanation during the second administration and 14% provided a different explanation altogether. On the other hand, only 28% of the individuals who worked collaboratively to generate an argument for the core ice-melting blocks problem produced an argument on

TABLE 2
The Proportion of Students in the Collaborative and Individual Argumentation Conditions That Used the Same, Similar, of Different Content During the Second Administration of the Ice-Melting Blocks Problem

	Content of the Explanation			Content of the Evidence		
Condition	Same	Similar	Different	Same	Similar	Different
Collaborative	0.28	0.37	0.35	0.55	0.13	0.32
Individual	0.54	0.32	0.14	0.64	0.25	0.11

the second (individual) administration that consisted of an explanation that had the same content as the explanation chosen by their triad (on the first administration). Thirty-seven percent of these individuals produced an argument that contained a similar explanation to the one used by their triad and 35% used a different explanation when working on their own. A chi-square goodness-of-fit test confirmed that there was a significant difference between the two treatment conditions in terms of these patterns of performance, χ^2 (2) = 18.45, $p < .001$.

A similar pattern is seen in the students' use of evidence. As shown in Table 2, 64% of the individuals who first worked individually to generate an argument for the ice-melting blocks problem produced an argument during the second administration that contained the same evidence as they used during the first administration of the problem. Twenty-five percent produced an argument that contained similar evidence during the second administration and 11% used different evidence altogether. In contrast, 55% of the individuals who first produced an argument for the ice-melting blocks problem as part of a triad generated an argument with evidence that had the same pieces of information as the evidence used by their triad. Thirteen percent of these individuals generated an argument that contained evidence similar to the evidence used by their triad and 32% used almost entirely different evidence when working on their own. A chi-square goodness-of-fit test once again confirmed that there was a significant difference between the two treatment conditions in terms of these patterns of performance, χ^2 (2) = 11.35, $p < .01$.

Although the results of this analysis indicate that a greater proportion of the students in the collaborative argumentation condition produced an argument with similar or different content when working on their own, it does not indicate whether these arguments were better, worse, or equivalent in terms of quality. To examine this issue, the quality of the arguments generated by each individual during the second administration of the ice-melting blocks problem was classified as either (a) superior, (b) equivalent, or (c) inferior in relation to quality of the arguments produced during the original administration of the ice-melting block problem. A superior performance was defined as a score that was greater than one standard deviation above the original score. An equivalent performance was defined as a score that was less than or equal to one standard deviation above or below original score. An inferior performance was defined as a score that was greater than one standard deviation below the original score. The proportion of students who fell into each category in the individual condition was then compared with the proportion of students who fell into each category in the collaborative condition. These data are provided in Table 3.

As can be seen in Table 3, 86% of the individuals who worked to generate an argument for the ice-melting blocks problem on their own produced an argument of equal quality when asked to solve the problem for the second time. Twelve percent produced an inferior

TABLE 3
The Proportion of Students in the Collaborative and Individual Argumentation Conditions Whose Score on the Second Administration of the Ice-Melting Blocks Problem Was Superior, Equivalent, and Inferior in Comparison to Their Original Performance on the Problem

	Overall Quality of the Argument		
Condition	Superior	Equivalent	Inferior
Collaborative	0.13	0.52	0.35
Individual	0.02	0.86	0.12

argument during the second administration and 2% showed improvement. On the other hand, 52% of the individuals who worked to generate an argument for the ice-melting blocks problem as part of a group produced an argument that was equivalent in quality to the argument produced by their triad. Thirty-five percent of these individuals produced an inferior argument when compared to the one produced by their triad, and 13% produced a better argument when working on their own. A chi-square goodness-of-fit test confirmed that there was a significant difference between the two treatment conditions in terms of these patterns of performance, χ^2 (2) = 22.27, $p < .001$.

Three patterns in these data warrant further discussion. First, as suggested by the challenges documented in the literature, group members do not universally adopt and internalize the group argument even when group members appear to reach consensus and work together to produce a single argument that articulates and justifies an explanation for a natural phenomenon. The proportion of students in the collaborative argumentation condition that constructed an argument with a different explanation or evidence provides clear evidence for this point. Second, while internalization is not complete or universal, however, the results of this analysis do indicate that a substantial number of individuals in the collaborative condition did incorporate their group's argument (or at least elements of it) as their own during the second administration of the problem. Therefore, some degree of adoption and internalization does occur. As illustrated in Table 2, 65% of the students who worked collaboratively used the same or a similar explanation as their group and 68% used the same or similar evidence. This indicates that many of the groups achieved a level of shared understanding or conceptual convergence (Rochelle, 1992) that is rarely seen during collaborative work (e.g., Forman, 1992; Moje & Shepardson, 1998; Southerland et al., 2005; Webb et al., 1995) Third, a sizable proportion of the students in the collaborative condition (13%) produced a better argument than their group did when working on their own compared to the 2% of students in the individual argumentation condition who produced an argument that was superior to their original argument during the second administration of the problem.

Clarifying the underlying reasons for these patterns will require further research. Certain factors and hypotheses, however, suggest themselves as initial candidates for exploration. For example, the individuals who did not adopt their group's argument might have taken a passive role (Webb, 1989) or the interactions that took place within some groups potentially did not stimulate conceptual change for some individuals (Amigues, 1988; Phelps & Damon, 1989; Webb & Palincsar, 1996). These factors would explain why some students did not adopt the arguments produced by their groups and why some students later produced inferior arguments on their own.

The task appeared, however, to foster a high level of conceptual convergence within some groups as evidenced by the percentage of students in groups that adopted all or part of their group's argument. We suspect that this was due, in part, to the kinds or argumentative acts that were required to complete the task and the types of cognitive processes that were triggered during these acts. For example, to evaluate the alternative explanations provided at the beginning of the task, students had to resolve differing perspectives through discussion, explain their thinking about a phenomenon, provide or receive critiques, and listen to reasoning of others. These cognitive processes have been shown to promote learning during collaborative work in other studies (Amigues, 1988; King, 1990; Linn & Eylon, 2006; Phelps & Damon, 1989; Webb & Palincsar, 1996). To substantiate this claim, however, targeted research that looks specifically for the occurrence of these types of cognitive processes in groups that did and did not achieve high levels of conceptual convergence during collaborative scientific argumentation is needed.

Finally, the fact that 13% of the students that worked in a group later produced a superior argument suggests that either the input of these students was not included in the creation of their groups' arguments or these students benefited from the collaborative exercise and were able to build further on their ideas in the subsequent phases of the activity. To help shed some light on this issue, we identified which group these individuals worked in and then examined the performance of these groups. This simple analysis indicated that 82% of the individuals that produced a superior argument when working on their own worked in different groups and 73% of them worked in a low performing group. These data suggest that these individuals produced superior arguments when working on their own because the interactions that took place within their groups either prevented these individuals from contributing or the other group members did not value these individuals' input. This explanation is also supported by the analysis that compared the interactions that took place in low and high performing groups (Sampson, 2007; Sampson and Clark, 2008) that we described in the discussion of the first research question. That analysis, once again, indicated that individuals in the lower performing groups introduced fewer content-related ideas into conversation and the members of these groups were quick to accept or reject ideas without critical discussion. Taken together, these observations provide some evidence, albeit indirect, that the input of these students was not included in the creation of their groups' arguments. This type of analysis, however, does not rule out the possibility that these students might also have been stimulated by the collaborative activity and subsequently produced superior arguments as a result of additional processing on their own at a later time.

In summary, the results of this study indicate that a high proportion of the students in the collaborative argumentation conditions adopted and internalized at least some component of their groups' argument. The rate of adoption for these students, however, was far from universal, and a substantial proportion of these students produced arguments that were better than their group at a later time. This suggests that the nature of task promoted conceptual convergence within most groups, but the interactions that took place within some groups constrained this process and prevented some groups from working at their fullest potential. As noted in the discussion of the first research question, it is important to interpret these findings in light of the sample and nature of the intervention. It is highly unlikely that students who lack familiarity and expertise with collaboration and the goals, processes, and affordances of scientific argumentation will be able to reap all the potential benefits of this type of knowledge building and validating process. As students gain more experience, familiarity, and expertise with collaborative scientific argumentation, we might see even higher quality group products and levels of adoption of group ideas by the individual group members. The findings from the current study therefore represent a baseline outcome that illustrates students' preexisting resources and skills related to collaborative argumentation, which teachers and researchers might further build upon and leverage through additional experiences and targeted instruction. These findings also suggest potential candidates for additional research that focuses on the improvement of group outcomes through the refinement of activity structures and group interactions.

Do Students Who Engage in Argumentation With Others Demonstrate Superior Performance on the Mastery and Transfer Task Than Students Who Engage in Argumentation Alone?

To assess individual learning outcomes as a function of an individual's ability to generate an argument that articulates and justifies an explanation for the mastery ice-melting blocks problem and the ability to transfer and apply this knowledge when asked to generate an

TABLE 4
The Means and Standard Deviation for Performance on the Mastery Problem and the Transfer Problem as a Function of Condition and Gender

		Mastery Problem		Transfer Problem	
Condition	Gender	*M*	*SD*	*M*	*SD*
Collaborative	Male	6.49	1.99	5.77	1.89
	Female	6.93	2.04	6.27	1.83
	Overall	6.73	2.02	6.04	1.87
Individual	Male	6.24	1.62	5.34	1.68
	Female	5.75	2.09	4.76	1.97
	Overall	5.92	1.95	4.96	1.86

argument that articulates and justifies an explanation for a conceptually similar phenomenon, individual student scores were analyzed using a $2 \times 2 \times 2$ MANOVA design. This design incorporated the argumentation condition (collaborative vs. individuals) and gender as the independent variables and the mastery (ice-melting blocks) and transfer (why do objects feel different) questions as the dependent measures. It is important to note here that this analysis does not compare performance on the core version of the ice-melting blocks problem with performance on the mastery problem or performance on the transfer problem. During the first problem-solving session, students who worked in triads produced only one argument and hence received a single score. Conducting a repeated measures analysis would therefore be inappropriate because of nonindependence between the scores. Thus, this analysis only compares the performance on the mastery and transfer problems as a function of condition.

The results of the MANOVA indicate that there were significant differences among the two grouping conditions on the dependent measures, Wilks's $\Lambda = 0.93$, $F\ (2, 163) = 5.86$, $p < .01$, multivariate $\eta^2 = .07$. There was no significant main effect for gender, Wilks's $\Lambda = 1.00$, $F\ (2, 163) = 0.01$, $p = .99$, or an interaction effect between condition and gender, Wilks's Λ 0.98, $F\ (2, 163) = 2.01$, $p = .14$. The strength of the relationship between condition and the mastery and transfer scores, as assessed by the multivariate η^2 associated with Wilks's Λ, was moderate, with 7% of the multivariate variance of the dependent variables contributed to the condition factor. Table 4 contains the means and standard deviations on each dependent variable for both conditions as a function of gender.

An ANOVA on each dependent variable was then conducted as follow-up tests to the MANOVA. Using the Bonferroni method to control for Type I error, each ANOVA was tested at the 0.025 level (0.05/2). The ANOVA on the mastery problem scores was significant, $F(1, 164) = 5.21$, $p = .02$, $d = 0.41$. Individuals in the collaborative argumentation condition ($M = 6.73$, $SD = 2.02$) scored significantly higher than the individuals assigned to the individual argumentation condition ($M = 5.92$, $SD = 1.95$). The ANOVA on the transfer problem scores was also significant, $F(1, 164) = 10.59$, $p < .01$, $d = 0.57$. Individuals in the collaborative argumentation condition ($M = 6.04$, $SD = 1.87$) scored significantly higher than the individuals assigned to the individual argumentation condition ($M = 4.96$, $SD = 1.86$). The effect sizes, as measured by Cohen's d, indicate a moderate magnitude of difference. See Figure 3.

Overall, this comparison of the arguments produced by each individual for the mastery and transfer problems indicates that students in the collaborative condition produced superior arguments on their own after collaboration. This observation suggests that collaboration can help students learn more *from* (i.e., content) and *about* (i.e., practices) scientific

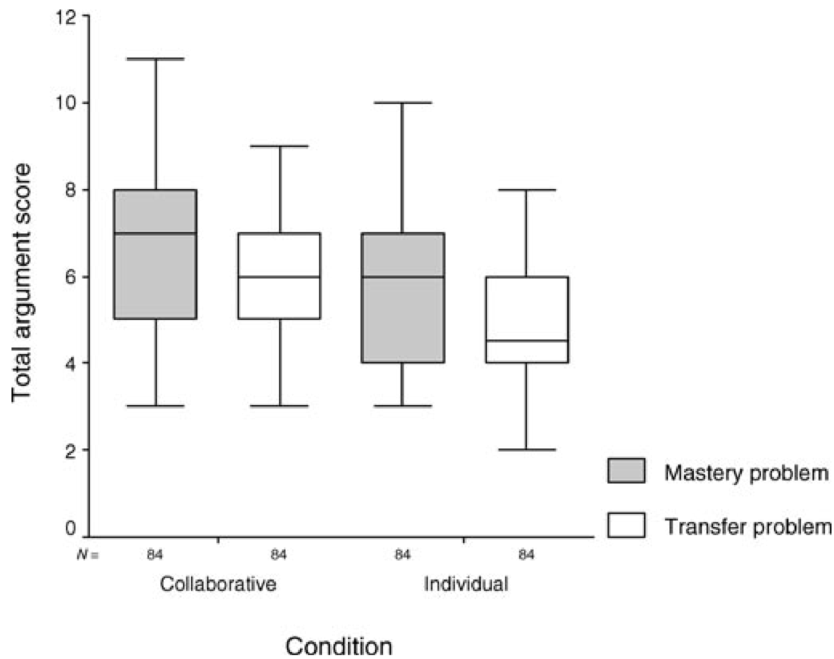


Figure 3. Individual performance on the mastery (ice-melting blocks) problem and the transfer (why do objects feel different) problem as a function of treatment condition.

argumentation in this context. This is an interesting and important result given the previous discussion of the first two questions. The value of collaboration during scientific argumentation for individual learning within the context of the current study seems to depend less on ideas becoming mutual knowledge (Gibbs & Mueller, 1990) or common ground within the group (Clark, 1996) and more on opportunities for individuals to "bounce" ideas off others, listen to other viewpoints, and think about the problem or what counts as a good argument in new ways or in a more deliberative manner. Independent of the mechanism, however, the finding that students in the collaborative condition produce superior arguments on mastery and transfer tasks is especially interesting given the relatively low "cost" of the intervention in terms of student preparation or experience with scientific argumentation. The simple act of grouping the students together, even with minimal instruction and past experience with scientific argumentation and collaboration, resulted in greater learning in the same amount of time.

SUMMARY AND CONCLUSIONS

Although students bring a number of productive argumentative resources with them to the classroom from previous experiences and everyday life, students often struggle with many of the nuances of scientific argumentation. Students often struggle to make sense of puzzling phenomena based on data, articulating explanations, supporting a viewpoint, and challenging the validity of ideas in a manner that reflects the nature of the argumentation that takes place within the scientific community. A number of science educators have therefore suggested that students should collaborate with each other as they work to make scientific argumentation more productive or to help distribute the cognitive workload. Few studies, however, have examined the effect of collaboration on argumentation outcomes or individual learning. This study addressed three questions that stem from this gap in the literature. First, do groups of students craft higher quality arguments than students working alone? Second, to what extent do students who work in groups adopt the arguments generated by their group? Third, do students who work in groups generate better arguments on mastery questions and transfer questions than individuals working alone?

In terms of the first question, the results of this study suggest that an opportunity to collaborate with others during this intervention did not have a substantial impact on overall argument quality. The triads, on average, produced arguments that were of higher quality than arguments produced by a group of the lowest performing individuals. However, the arguments of the triads were only equivalent in quality, on average, to the arguments produced by the second highest (middle) performing individuals in the study and inferior to the arguments produced by the highest performing individuals. These results suggest that collaboration did not have the impact that was expected given the extensive literature that suggests that collaborative effort can and should result in a product that exceeds what is possible by an individual working alone (e.g., Andriessen, Erkens, et al., 2003; Mason, 1998; Rochelle, 1992; Scardamalia & Bereiter, 1994). This may have resulted from the minimal instruction provided to the students, from the nature and structure of the task itself, or the nature of the interactions that took place between individuals, ideas, and materials in some of the groups.

In terms of the second question, the results of this study indicate that a substantial proportion of the students in the collaborative argumentation condition adopted their groups' argument, or at least elements of the argument, although the adoption of group products was clearly not universal. These observations suggest that the nature or structure of the task promoted a high level of conceptual convergence within many of the groups. This potentially resulted from the cognitive processes (e.g., evaluating alternatives, resolving different perspectives through discussion, explaining one's thinking about a phenomenon, and providing or receiving critiques) triggered during the task. The results of this study also indicate that 13% of students in the collaborative argumentation condition produced better arguments than their group when working on their own, however, and a majority of these students worked in low performing groups. These observations underscore the idea that group interactions can and do constrain group meaning making and/or the nature of the product produced by a group. It seems, however, that more research that examines innovative ways to improve group function to better engage all group members in the process of collaborative scientific argumentation will be needed to clarify the underlying reasons for these observations.

In terms of the third question, the results of this study indicate that the students who worked in a group produced significantly better arguments (with moderate effect sizes) than the students who worked alone on both the mastery and transfer tasks. This finding suggests that collaboration improves what students learn both *from* and *about* scientific argumentation when students engage in a task that requires the evaluation of alternative explanations for a discrepant event and then the generation of an argument that provides and justifies an explanation for the phenomenon under investigation. In light of the previous conclusions, these results suggest that value of the collaborative argumentation activity for individual learning during this type of task comes from opportunities to "bounce" ideas off others, listen to other viewpoints, and think about the problem or what counts as a good argument in new ways or in a more deliberative manner rather than group performance. It also suggests that the simple act of grouping students together in this task structure for scientific argumentation will result in greater individual learning without requiring extra time or teacher expertise.

POTENTIAL LIMITATIONS OF THE STUDY

There are five main limitations to this work. First, although every effort was made to equate the two conditions with respect to the learning environment, some differences remained. One difference was the presence of a video camera for those individuals in the group argumentation condition. Although it is possible that the presence of a video camera may have influenced student behavior, it is difficult to predict in which direction. For example, students may have been more focused on the task because they knew that they were being recorded or the camera could have distracted them. Informal observation, however, suggested that students in both conditions were very attentive and were highly engaged as they worked.

Second, the need for experimental control over variables such as the types of explanations students evaluated and the data they had available resulted in a very specific task. As discussed earlier, there are a number of different ways to promote argumentation inside the classroom and the task used in this study represents but one of many. It is therefore important to keep in mind that the results may have been different, if the nature of the task were changed. For example, an opportunity to collaborate with others could be far more valuable during a task that required students to develop their own explanations for the phenomenon under investigation or to gather their own data to evaluate or support an explanation. The results of the current study therefore call for further replication with different types of tasks to support broader generalization and specification.

The third limitation involves the selection of groups. While substantial work has focused on optimizing grouping in terms of achievement levels, the current study did not sort students into groups based on achievement levels. Integrating research on grouping and achievement level into the group formation process might therefore enhance the efficacy of the approach.

The fourth potential limitation of this study is that the participants in this study, as noted earlier, did not collaborate with each other or engage in argumentation on a regular basis. Thus, the "unwritten rules" (Lemke, 1990) of these classrooms may have had a negative impact on the ways students assigned to the collaborative argumentation condition worked with each other during the intervention. For example, Kuhn and Reiser (2006) suggest that most students have no need or motivation to engage with their classmates' ideas in productive ways because this type of activity is of little value inside most classrooms. As a result, students who work with others often look to the most "capable member" for the desired answer (Hatano & Inagaki, 1991) or breakup tasks into smaller pieces that individuals can accomplish on their own (Cohen, 1994; Eichinger, Anderson, Palincsar, & David, 1991) because these strategies are often more efficient in the classroom. Although these types of strategies can be valuable during some types of group work, these strategies can act as a barrier to generating a high-quality argument that articulates and justifies an explanation for a discrepant event. Therefore, if the participants had more experience with tasks that require argumentation or genuine collaboration, the effect of collaboration on argumentation outcomes might have been more substantial. Nevertheless, these findings indicate that science educators interested in promoting and supporting optimal productive argumentation in the classroom will need to do more than simply group students together and tell them to argue. Science educators will also need to help students learn how to interact with other students, ideas, and materials in a more productive manner.

The fifth limitation also focuses on the students involved. While the students in this study did not have much formal classroom experience or instruction about the goals and processes of scientific argumentation, these students were drawn from six chemistry classes in a public high school and therefore represent a fairly high achieving group of older students as defined by their enrollment in an elective science course. This, however, is only a limitation to the degree that it potentially limits the generalizability of the conclusions in terms of the age and prior academic success of the students involved. Future studies could replicate the activities of this study with younger students to extend generalizability in this regard. Such studies would also provide further insight into the argumentative and collaborative resources that students bring to these activities and the development of these resources over time.

IMPLICATIONS AND FINAL THOUGHTS

The current literature outlines a number of ways to structure classroom instruction so that students have the opportunity to propose, support, critique, and revise ideas. For example, students can evaluate the validity or acceptability of alternative explanations based on available information or they can discuss the validity or acceptability of the process, context, or product of an inquiry. Clearly, these activities engage students in the processes of scientific argumentation inside the classroom. Research now needs to better understand what and how students learn *from* or *about* scientific argumentation as part of these activities. The results of this study have a number of implications for this type of work.

First, the students in this study constructed better arguments, in terms of content and justification, on their own after collaborating with others even though the students in this study had minimal prior formal experience or instruction in the nuances of scientific argumentation. This suggests that an opportunity to collaborate with others can help students learn more from (i.e., content) and about scientific argumentation (i.e., practices). It also suggests that students bring sufficient cognitive and monitoring resources to the classroom to learn during this type of activity despite the well-documented challenges students encounter with the nuances of scientific argumentation. Students would potentially demonstrate enhanced learning outcomes, however, with additional prior experience and instruction in collaborative scientific argumentation. As noted earlier, these results therefore represent a baseline rather than an upper limit.

Second, this study provides evidence that the benefits of collaboration are not universal. Some groups, as noted earlier, seem to resolve differences in understanding and negotiate issues of individual and collective action to develop a quality group product through scientific argumentation more effectively than others. This indicates that group outcomes are not simply the sum of individual abilities—group interaction processes clearly influence group outcomes. Although research has focused on understanding the factors that can constrain group productivity in other contexts, less is known about the interactions that influence group performance during scientific argumentation. This line of research holds great promise in terms of supporting and promoting more productive scientific argumentation inside the classroom.

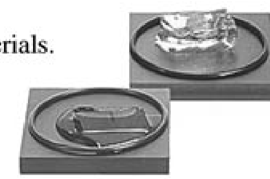
Finally, the results of this study indicate that some individuals appear to benefit more than others from collaborative scientific argumentation. Some individuals adopt the group outcomes, some only adopt elements, and some do not adopt any of the ideas of the group as their own. Some students perform more successfully working on their own and some less successfully. How individuals interact and the decisions they make while engaged in collaborative scientific argumentation appear to influence both group meaning making and individual meaning making. Taken together, these observations suggest the need for further targeted research exploring how individuals influence group processes and how the ideas and processes of the group influence the understandings of the individuals.

APPENDIX A

The Ice-Melting Blocks Problem and the Why Do Objects Feel Different Problem

On the table in front of you are two black blocks, A and B.

These blocks look the same but they are made of different materials.



Place an ice cube on each block and watch how long it takes for the ice cube to melt on each of these blocks.

Use the data provided to you in order to answer the following research question: **Why does the ice melt faster on block A?**

a) What is your explanation?
b) How do you know? In the space below, defend your explanation with appropriate *evidence* and *reasoning*.

Examine the following data table. It provides information about four different objects that have been sitting in the same room for 24 hours. The thermostat on the wall is set at 23°C.

Object	Mass (g)	Density (g/mL)	Temperature (°C)	How It Feels	Thermal Conductivity	Temperature Change When Placed in a 65°C Oven for 15 Minutes (°C)
Metal Spoon	48	7.4	23.0	Cold	High	+26
Pencil	20	0.7	23.1	Warm	Low	+17
Empty Glass	64	2.6	23.0	Cool	Medium	+21
Styrofoam Cup	34	0.01	23.0	Warm	Low	+14
Penny	5	8.9	22.9	Cold	High	+34

Use this information to answer the following research question: **Why do some objects feel hotter or colder than others even though they have been sitting in the same room for long periods of time?**

a) What is your explanation?
b) How do you know? In the space below, defend your explanation with appropriate *evidence* and *reasoning*.

APPENDIX B

Sample Items From the Thermodynamics Content Knowledge Test

An empty plastic coffee cup and stainless steel coffee cup have been in sitting Ms. D's classroom all day. You look at the thermostat in the classroom and it says that it is 23°C in the room. You then measure the temperature of the two coffee cups using a thermometer.

2. The temperature of the stainless steel coffee cup will be. . .

   a. much cooler than the room ($<$19°C).
   b. slightly cooler than the room (21–22°C).
   c. the same temperature as the room (23°C).
   d. slightly warmer than the room (24–25°C).
   e. much hotter than the room ($>$27°C).

   Why?

   a. The metal coffee cup does not produce its own heat.
   b. Metal is a good conductor of heat.
   c. Metal takes a long time to reach the temperature of its surroundings.
   d. Metal absorbs cold from its surroundings.
   e. Metal feels colder than other objects.

3. The temperature of the plastic coffee cup will be. . .

   a. Much cooler than the room ($<$19°C).
   b. Slightly cooler than the room (21–22°C).
   c. The same temperature as the room (23°C).
   d. Slightly warmer than the room (24–25°C).
   e. Much hotter than the room ($>$27°C).

   Why?

   a. The plastic coffee cup feels warmer than the metal cup.
   b. The plastic coffee cup is not producing its own heat energy.
   c. Plastic is a good conductor of heat.
   d. Plastic needs a lot of time to reach the temperature of its surroundings.
   e. Plastics are good insulators.

Predict what will happen if you use the materials described below as either a storage container or a wrap.

13. Aluminum foil is:

   a. Good at keeping hot things hot. Bad at keeping cold things cold.
   b. Bad at keeping hot things hot. Good at keeping cold things cold.
   c. Good at keeping hot things hot. Good at keeping cold things cold.
   d. Bad at keeping hot things hot. Bad at keeping cold things cold.

   Why?

   a. Metal absorbs cold from its surroundings.
   b. Heat travels through it quickly.
   c. Metal warms things up that are cold.
   d. Metal attracts and absorbs heat.
   e. Heat cannot travel through metal.

14. Wool is:

a. Good at keeping hot things hot. Bad at keeping cold things cold.
b. Bad at keeping hot things hot. Good at keeping cold things cold.
c. Good at keeping hot things hot. Good at keeping cold things cold.
d. Bad at keeping hot things hot. Bad at keeping cold things cold.

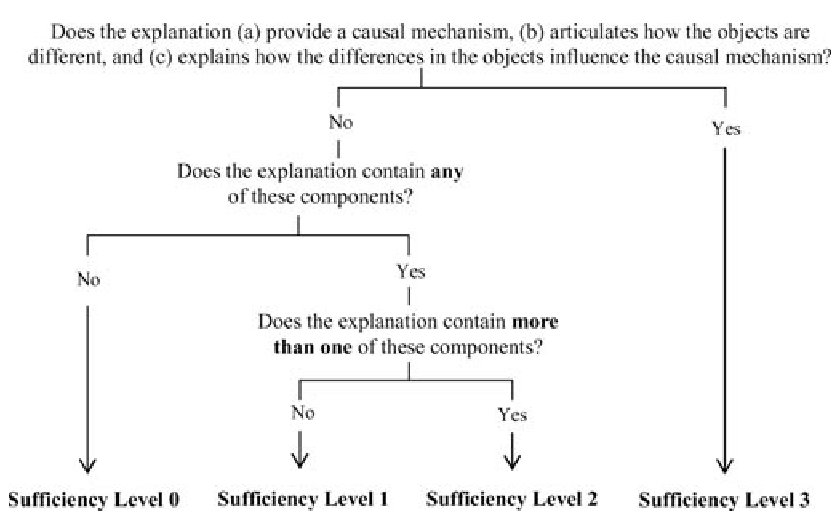
Why?

a. Wool heats up slowly.
b. Heat travels through wool slowly.
c. Wool warms things up that are cold.
d. Wool is a thick material.
e. Wool has holes in it.

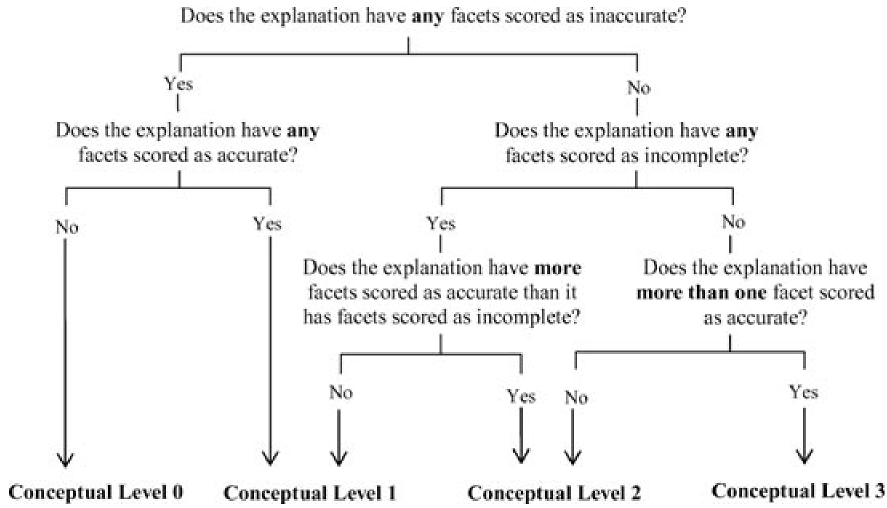
APPENDIX C

Examples of Keys Used to Score the Arguments

Sufficiency of the Explanation



Conceptual Quality of the Explanation



REFERENCES

Abell, S. K., Anderson, G., & Chezem, J. (2000). Science as argument and explanation: Exploring concepts of sound in third grade. In J. Minstrell & E. H. Van Zee (Eds.), Inquiry into inquiry learning and teaching in science (pp. 100–119). Washington, DC: American Association for the Advancement of Science.

Aikenhead, G. (2004). Science-based occupations and the science curriculum: Concepts of evidence. Science Education, 89(2), 242–275.

Alexopoulou, E., & Driver, R. (1996). Small-group discussion in physics: Peer interaction modes in pairs and fours. Journal of Research in Science Teaching, 33(10), 1099–1114.

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

Amigues, R. (1988). Peer interactions in solving physics problems: Sociocognitive confrontation and metacognitive aspects. Journal of Experimental Child Psychology, 45(1), 141–158.

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.

Andriessen, J., Erkens, G., Van de Laak, C., Peters, N., & Coirier, P. (2003). Argumentation as negotiation in electronic collaborative writing. In J. Andriessen, M. Baker, & D. Suthers (Eds.), Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments (pp. 79–115). Dordrecht, The Netherlands: Kluwer.

Azmitia, M. (1988). Peer interaction and problem solving: When are two heads better than one? Child Development, 59(1), 87–96.

Bakeman, R., & Gottman, J. (1986). Observing interaction: An introduction to sequential analysis. Cambridge, England: Cambridge, University Press.

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). Problem solving in video-based microworlds: Collaboration and individual outcomes of high-achieving sixth grade students. Journal of Educational Psychology, 92(2), 391–398.

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.

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

Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). "An experiment is when you try it and see if it works": A study of grade 7 students' understanding of the construction of scientific knowledge. International Journal of Science Education, 11(Special Issue), 514–529.

Carey, S., & Smith, C. L. (1993). On understanding the nature of scientific knowledge. Educational Psychologist, 28(3), 235–251.
Cartier, J. L., & Stewart, J. (2000). Teaching the nature of inquiry: Furthur developments in a high school genetics curriculum. Science and Education, 9(3), 247–267.
Chi, M. T. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing metal models. In R. Glaser (Ed.), Advances in instructional psychology. Mahwah, NJ: Erlbaum.
Chi, M. T., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. Cognitive Science, 13, 145–182.
Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. Review of Educational Research, 63, 1–49.
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., & Linn, M. C. (2003). Designing for knowledge integration: The impact of instructional time. Journal of the Learning Sciences, 12(4), 451–494.
Clark, D., & Sampson, V. (2005). Analyzing the quality of argumentation supported by personally-seeded discussions. Paper presented at the Computer Supported Collaborative Learning Conference, Taipei, Taiwan.
Clark, D., & Sampson, V. (2006a). 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.
Clark, D., & Sampson, V. (2006b). Personally-seeded discussions to scaffold online argumentation. International Journal of Science Education, 29(3), 253–277.
Clark, H. H. (1996). Using language. Cambridge, England: Cambridge University Press.
Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. Review of Educational Research, 64, 1–35.
Coleman, E. B. (1998). Using explanatory knowledge during collaborative problem solving in science. Journal of the Learning Sciences, 7(3,4), 387–427.
deVries, E., Lund, K., & Baker, M. (2002). Computer-mediated epistemic dialogue: Explanation and argumentation as vehicles for understanding scientific notions. Journal of the Learning Sciences, 11(1), 63–103.
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. (2000). Making the nature of science explicit. In R. Millar, J. Leach, & J. Osborne (Eds.), Improving science education: The contribution of research. Philadelphia: Open University Press.
Duschl, R. (2007). Quality argumentation and epistemic criteria. In S. Erduran & M. Jimenez-Aleixandre (Eds.), Argumentation in science education: Perspectives from classroom-based research (pp. 159–175). Dordrecht, The Netherlands: Springer Academic Publishers.
Duschl, R. A. (1990). Restructuring science education: The importance of theories and their development. New York: Teachers College Press.
Duschl, R. A., Ellenbogan, E., & Erduran, S. (1999). Promoting argumentation in middle school classrooms: A project SEPIA evaluation. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Boston, MA.
Duschl, R. A., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. Studies in Science Education, 38, 39–72.
Eichinger, D., Anderson, C. W., Palincsar, A. S., & David, Y. M. (1991). An illustration of the roles of content knowledge, scientific argument, and social norm in collaborative problem solving. Paper presented at the annual meeting of the American Educational Research Association, April 1991, Chicago, IL.
Eisenberg, A., & Garvey, C. (1981). Children's use of verbal strategies in resolving conflict. Discourse Processes, 4, 149–170.
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.
Forman, E. (1992). Discourse, intersubjectivity and the development of peer collaboration: A Vygotskian approach. In L. T. Winegar & J. Valsinier (Eds.), Children's development within social context (pp. 143–159). Hillsdale, NJ: Erlbaum.

Forman, E. A., Larreamendy-Joerns, J., Stein, M. K., & Brown, C. A. (1998). "You're going to want to find out which and prove it": Collective argumentation in a mathematics classroom. Learning and Instruction, 8(6), 527 – 548.

Forman, E., & Cazden, C. (1985). Exploring Vygotskian perspectives in education: The cognitive value of peer education. New York: Cambridge University Press.

Gallas, K. (1995). Talking their way into science: Hearing children's questions and theories, responding with curricula. New York: Teachers College Press.

Gibbs, R., & Mueller, R. (1990). Conversation as coordinated, cooperative interaction. In S. Robertson, W. Zachaary, & J. Black (Eds.), Cognition, computing, and cooperation. Norwood, NJ: Ablex.

Goldman, S., Cosden, M., & Hine, M. (1992). Working alone and working together: Individual differences in the effects of collaboration on learning handicapped students' writing. Learning & Individual Differences, 4(4), 369 – 393.

Goldman, S. R., Duschl, R. A., Ellenbogen, K., Williams, S. M., & Tzou, C. (2002). Science inquiry in a digital age: Possibilities for making thinking visible. In V. Oostendorp (Ed.), Cognition in a digital age (pp. 253 – 281). Mahwah, NJ: Erlbaum.

Hammer, D. (2000). Student resources for learning introductory physics. American Journal of Physics, Supplement 68(7), S52 – S59.

Hammer, D., & Elby, A. (2003). Tapping epistemological resources for learning physics. Journal of the Learning Sciences, 12(1), 53 – 90.

Hatano, G., & Inagaki, K. (1991). Sharing cognition through collective comprehension activity. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), Perspective on socially shared cognition (pp. 331 – 348). Washington, DC: American Psychological Association.

Hill, G. W. (1982). Group versus individual performance: Are N + 1 heads better than one? Psychological Bulletin, 91, 517 – 539.

Hogan, K. (2000). Exploring a process view of students' knowledge about the nature of science. Science Education, 84, 51 – 70.

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.

Hutchins, E. (1995). Cognition in the wild. Cambridge, MA: MIT Press.

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.

Kelly, G. J., & Chen, C. (1999). The sound of music: Constructing science as a sociocultural practice through oral and written discourse. Journal of Research in Science Teaching, 36(8), 883 – 915.

Kelly, G. J., Druker, S., & Chen, C. (1998). Students' reasoning about electricity: Combining performance assessments with argumentation analysis. International Journal of Science Education, 20(7), 849 – 871.

Kenyon, L. O., & Reiser, B. J. (2005). Students' epistemologies of science and their influence on inquiry practice. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Dallas, TX.

King, A. (1990). Facilitating elaborative learning in the classroom through reciprocal questioning. American Educational Research Journal, 27, 664 – 687.

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, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. Science Education, 77(3), 319 – 337.

Kuhn, D., Garcia-Mila, M., Zohar, A., & Anderson, C. (1995). Strategies of knowledge acquisition. Monographs of the Society for Research in Child Development 60 (4, Serial No. 245).

Kuhn, D., & Udell, W. (2003). The development of argument skills. Child Development, 74(5), 1245 – 1260.

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.

Laughlin, P., Hatch, E., Silver, J., & Boh, L. (2006). Groups perform better than individuals on letters-to-numbers problems: Effects of group size. Journal of Personality and social Psychology, 90(4), 644 – 651.

Lawson, A. (2003). The nature and development of hypothetico-predictive argumentation with implications for science teaching. International Journal of Science Education, 25(11), 1387 – 1408.

Lemke, J. (1990). Talking science: Language, learning, and values. Norwood, NJ: Ablex.
Lewis, E. L. (1996). Conceptual change among middle school students studying elementary thermodynamics. Journal of Science Education and Technology, 5(1), 3–31.
Lewis, E. L., & Linn, M. C. (1994). Heat energy and temperature concepts of adolescents, adults, and experts: Implications for curricular improvements. Journal of Research in Science Teaching, 31(6), 657–677.
Linn, M. C., Eylon, B., & Davis, E. A. (2004). The knowledge integration perspective on learning. In M. C. Linn, E. A. Davis, & P. Bell (Eds.), Internet environments for science education (pp. 29–46). Mahwah, NJ: Erlbaum.
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.
Linn, M. C., & Hsi, S. (2000). Computers, teachers, peers: Science learning partners. Mahwah, NJ: Erlbaum.
Lizotte, D. J., Harris, C. J., McNeill, K. L., Marx, R. W., & Krajcik, J. (2003). Usable assessments aligned with curriculum materials: Measuring explanation as scientific way of knowing. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
Lizotte, D. J., McNeill, K. L., & Krajcik, J. (2004). Teacher practices that support students' construction of scientific explanations in middle school classrooms. In Y. Kafai, W. Sandoval, N. Enyedy, A. Nixon, & F. Herrera (Eds.), Proceedings of the 6th International Conference of the Learning Sciences (pp. 310–317). Mahwah, NJ: Erlbaum.
Mason, L. (1998). Sharing cognition to construct scientific knowledge in school context: The role of oral and written discourse. Instructional Science, 26, 359–389.
May, D., Hammer, D., & Roy, P. (2006). Children's analogic reasoning in a 3rd grade science discussion. Science Education, 90(2), 316–330.
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.
Moje, E. B., & Shepardson, D. P. (1998). Social interactions and children's changing understanding of electric circuits: Exploring unequal power relations in "peer" learning groups. In B. Guzzetti & C. Hynd (Eds.), Perspectives on conceptual change (pp. 225–234). Mahwah, NJ: Erlbaum.
Monk, M., & Osborne, J. (1997). Placing the history and philosophy of science in the curriculum: A model for the development of pedagogy. Science Education, 81(4), 405–424.
National Research Council. (1996). National Science Education Standards. Washington, DC: National Academy Press.
National Research Council. (2000). Inquiry and the National Science Education Standards. Washington, DC: National Academy Press.
Newton, P., Driver, R., & Osborne, J. (1999). The place of argumentation in the pedagogy of school science. International Journal of Science Education, 21(5), 553–576.
Odom, A. L., & Barrow, L. H. (1995). The development and application of a two-tiered diagnostic test measuring college biology students' understanding of diffusion and osmosis following a course of instruction. Journal of Research in Science Teaching, 32, 45–61.
Ohlsson, S. (1992). The cognitive skill of theory articulation: A neglected aspect of science education? Science & Education, 1, 181–192.
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.
Perkins, D. N., Farady, M., & Bushy, B. (1991). Everyday reasoning and the roots of intelligence. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), Informal reasoning and education. Hillsdale, NJ: Erlbaum.
Phelps, E., & Damon, W. (1989). Problem solving with equals: Peer collaboration as a context for learning mathematics and spatial concepts. Journal of Educational Psychology, 81(4), 639–646.
Pontecorvo, C., & Girardet, H. (1993). Arguing and reasoning in understanding historical topics. Cognition and Instruction, 11(3–4), 365–395.
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.
Richmond, G., & Striley, J. (1996). Making meaning in the classroom: Social processes in small-group discourse and scientific knowledge building. Journal of Research in Science Teaching, 33(8), 839–858.

Rochelle, J. (1992). Learning by collaborating: Convergent conceptual change. Journal of the Learning Sciences, 2, 235–276.

Roth, K. J., Druker, S. L., Garnier, H., Lemmens, M., Chen, C., Kawanaka, T., et al. (2006). Teaching science in five countries: Results from the TIMSS 1999 video study. Washington, DC: National Center for Education Statistics.

Roth, W.-M., McGinn, M. K., Woszcyna, C., & Boutonne, S. (1999). Differential participation during science conversations: The interaction of focal artifacts, social configurations, and physical arrangements. Journal of the Learning Sciences, 8(3–4), 293–347.

Sadler, T. (2004). Informal reasoning regarding socioscientific issues: A critical review of the research. Journal of Research in Science Teaching, 41(5), 513–536.

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

Sampson, V. (2007). The effects of collaboration of argumentation outcomes. Unpublished Doctoral Dissertation, Arizona State University, Tempe, AZ.

Sampson, V., & Clark, D. (2008). A case study comparison of the scientific argumentation practices that occur in two high and two low performing groups. Manuscript submitted for publication.

Sandoval, W. A. (2003a). Conceptual and epistemic aspects of students' scientific explanations. Journal of the Learning Sciences, 12(1), 5–51.

Sandoval, W. A. (2003b). The inquiry paradox: Why doing science doesn't necessarily change ideas about science. In C. P. Constantinou & Z. C. Zacharis (Eds.), Proceedings of the 6th International Computer-Based Learning in Science Conference (pp. 825–834). Nicosia, Cyprus.

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.

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. Journal of the Learning Sciences, 3(3), 265–283.

Schauble, L., Glaser, R., Duschl, R., Schulze, S., & John, J. (1995). Students' understanding of the objectives and procedures of experimentation in the science classroom. Journal of the Learning Sciences, 4(2), 131–166.

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.

Schwarz, B., & Glassner, A. (2003). The blind and the paralytic: Supporting argumentation in everyday and scientific issues. In J. Andriessen, M. Baker, & D. Suthers (Eds.), Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments (pp. 227–260). Dordrecht, The Netherlands: Kluwer.

Settlage, J., & Odom, A. L. (1995, April 22–25). Natural selection conceptions assessment: Development of a two-tiered test "Understanding Biological Change." Paper presented at the annual meeting of the National Association for Research in Science Teaching, San Francisco, CA.

Siegel, H. (1989). The rationality of science, critical thinking and science education. Synthese, 80(1), 9–42.

Siegel, H. (1995). Why should educators care about argumentation? Informal Logic, 17(2), 159–176.

Simon, S., Erduran, S., & Osborne, J. (2006). Learning to teach argumentation: Research and development in the science classroom. International Journal of Science Education, 28(2 & 3), 235–260.

Southerland, S., Kittleson, J., Settlage, J., & Lanier, K. (2005). Individual and group meaning-making in an urban third grade classroom: Red fog, cold cans, and seeping vapor. Journal of Research in Science Teaching, 42(9), 1032–1061.

Stein, N. L., & Bernas, R. (1999). The early emergence of argumentative knowledge and skill. In G. Rijlaarsdam, E. Esperet, J. Andriessen, & P. Coirier (Eds.), Studies in writing: Vol. 5. Foundations of argumentative text processing. Amsterdam: University of Amsterdam Press.

Stein, N. L., & Miller, C. (1991). I win. . . you lose: The development of argumentative thinking. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), Informal reasoning and instruction. Hillsdale, NJ: Erlbaum.

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.

Suthers, D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. International Journal of Computer Supported Collaborative Learning, 1(3), 315–337.

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 3rd

International Conference on Intelligent Tutoring Systems (ITS '96). Montreal, Quebec, Canada: Springer-Verlag.

Toth, E., Suthers, D., & Lesgold, A. (2002). "Mapping to know": The effects of representational guidance and reflective assessment on scientific inquiry. Science Education, 86(2), 264–286.

Treagust, D. F. (1988). Development and use of a diagnostic test to evaluate students' misconceptions in science. International Journal of Science Education, 10, 159–169.

van Eemeren, F., Grootendorst, R., & Henkemans, A. F. (2002). Argumentation: Analysis, evaluation, presentation. Mahwah, NJ: Erlbaum.

Vellom, R. P., & Anderson, C. W. (1999). Reasoning about data in middle school science. Journal of Research in Science Teaching, 36(2), 179–199.

Webb, N. (1984). Sex difference in interaction and achievement in cooperative small groups. Journal of Educational Psychology, 76, 33–44.

Webb, N. (1989). Peer interaction and learning in small groups. International Journal of Educational Research, 13, 21–39.

Webb, N., & Palincsar, A. (1996). Group processes in the classroom. In D. Berliner & R. Calfee (Eds.), Handbook of educational psychology (3rd ed.). New York: Macmillian.

Webb, N., Troper, J., & Fall, R. (1995). Constructive activity and learning in collaborative groups. Journal of Educational Psychology, 87, 406–423.

Webb, N. M. (1985). Verbal interaction and learning in peer-directed groups. Theory Into Practice, 24, 32–39.

White, R., & Gunstone, R. (1992). Probing understanding. London: Falmer Press.

White, R. T. (1992). Implications of recent research on learning for curriculum and assessment. Journal of Curriculum Studies, 24(2), 153–164.

Yerrick, R. (2000). Lower track science students' argumentation and open inquiry instruction. Journal of Research in Science Teaching, 37(8), 807–838.

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

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. Journal of Research in Science Teaching, 39(1), 35–62.