Res Sci Educ (2011) 41:63–97
DOI 10.1007/s11165-009-9146-9

A Comparison of the Collaborative Scientific Argumentation Practices of Two High and Two Low Performing Groups

Victor Sampson · Douglas B. Clark

Published online: 27 November 2009

Abstract This qualitative study examines the interactions between individuals, ideas, and materials as two high and two low performing groups of students engaged in a process of collaborative scientific argumentation. To engage students in collaborative scientific argumentation the students were randomly assigned to small groups of three students each. Each triad was asked to critique six alternative explanations for a discrepant event and to produce a single written argument justifying the explanation they felt was most valid or acceptable. The two higher performing triads produced arguments that included a sufficient and accurate explanation that was well supported with appropriate evidence and reasoning while the two lower performing triads produced arguments that included an inaccurate explanation supported by inappropriate justification. A verbal analysis of the interactive processes that took place within these four triads identified five distinct differences in the ways these triads engaged in collaborative scientific argumentation that seemed to promote or constrain the development of high quality written arguments. These differences include (1) the number of unique ideas introduced into the conversation, (2) how individuals responded to these ideas, (3) how often individuals challenged ideas when discussing them, (4) the criteria individuals used to distinguish between ideas, and (5) how group members used the available corpus of data. The conclusions and implications of this study include recommendations for the design and revision of curriculum, the development of new instructional models and technology-enhanced learning environments, and areas for future research.

Keywords Argumentation · Argument · Chemistry · Collaboration · Verbal analysis

V. Sampson (✉)
School of Teacher Education, The Florida State University, 205 Stone Building, Tallahassee, FL 32306-4459, USA
e-mail: vsampson@fsu.edu

D. B. Clark
Vanderbilt University, Nashville, TN, USA

Introduction

Instructional tasks that engage students in scientific argumentation can help foster scientific literacy (Abell et al. 2000; Driver et al. 2000; Duschl and Osborne 2002; Erduran et al. 2004a; Kuhn 1993). Scientific argumentation requires individuals to gather and make sense of data, generate and articulate explanations for natural phenomena, justify explanations with appropriate evidence and reasoning, and critique the validity and legitimacy of one or more viewpoints. Current research indicates that when students engage in these types of activities on a regular basis they can learn science content (Bell and Linn 2000; Zohar and Nemet 2002), develop complex reasoning and critical thinking skills (Lawson 2003; Sadler 2004; Siegel 1995), understand how knowledge is generated and validated in science (Driver et al. 2000; Osborne et al. 2004), and improve their communication skills (Kuhn and Udell 2003). This literature, however, also suggests that these opportunities are rare for students because traditional approaches to science instruction do not promote or support student engagement in scientific argumentation.

It is therefore not surprising that a number of collaborative instructional approaches have been developed over the last decade to promote and support student engagement in scientific argumentation inside the classroom. For example, students can be placed into small groups and asked to evaluate the merits of alternative explanations for a given phenomenon (e.g., Bell and Linn 2000; Clark and Sampson 2006b; Osborne et al. 2004) or they can be encouraged to generate an argument and evaluate the arguments of other groups as part of the inquiry process (e.g., Jimenez-Aleixandre et al. 2000; Kelly and Chen 1999; Kuhn and Reiser 2005; Sandoval and Reiser 2004). The goals of these collaborative approaches often focus on promoting the development of a useful intellectual outcome (e.g., a more sophisticated understanding of a topic or of the processes involved in scientific inquiry) or a high quality product (e.g., a sufficient explanation or an argument that provides and justifies an explanation). Unfortunately, empirical research indicates that collaboration does not always result in a beneficial outcome for students. A number of studies, for example, suggest that simply grouping students together and telling them to generate an argument or to evaluate the merits of alternative explanations for a given phenomenon will often result in significant variation in group performance (Jimenez-Aleixandre et al. 2000; Kuhn and Reiser 2005; Sampson and Clark 2009) and some individuals appear to be hindered by working with others (Barron 2003; Hogan et al. 2000; Richmond and Striley 1996; Sampson and Clark 2009). It seems that some groups engage in collaborative scientific argumentation in a way that fosters a high quality group outcome (e.g., new ideas, a written artifact, or other intellectual product) and some groups do not.

The underlying reasons for this inter-group variation in performance remain unclear. Some researchers (e.g., Andriessen et al. 2003b; Boulter and Gilbert 1995; deVries et al. 2002; Hogan et al. 2000; Suthers and Hundhausen 2001; Veerman 2003) suggest that groups that co-construct knowledge as they work (e.g., by adding to, transforming, or building off each others' ideas) perform better than groups that do not. This focus is grounded in the idea that better group outcomes often result from the generation of novel ideas or ideas that no individual possessed prior to the collaboration. Other researchers (e.g., Clark and Sampson 2006b; Erduran et al. 2004b; Kuhn and Udell 2003; Kuhn and Reiser 2006; Osborne et al. 2004; Southerland et al. 2005) suggest that a high level of oppositional discourse between group members (e.g., challenging, critiquing, and evaluating the ideas of others) leads to a better group outcome. This perspective is based on the idea that intellectual rigor and constructive criticism helps people eliminate or revise invalid ideas during collaborative scientific argumentation. A number of other researchers

(Andriessen et al. 2003a; Sandoval and Reiser 2004; Schwarz and Glassner 2003) also suggest that students must adopt different goals (e.g., improving understanding rather than attempting to win, etc.) and strategies (e.g., being skeptical of new ideas, making ideas more problematic, etc.) in order to develop a better group outcome. Finally, some researchers (Cartier and Stewart 2000; Kuhn and Reiser 2006) recommend that students learn to use different criteria than they would normally use in other contexts (e.g., fit with evidence rather than fit with personal experiences, etc.) to distinguish between competing ideas. These last two perspectives stress the importance of students learning, using, and ultimately, relying on the same epistemic framework that scientists use (Duschl 2007) when they engage in collaborative scientific argumentation.

This broad range of recommendations underscores a lack of consensus in the literature about the type of interactions that can and should be cultivated within groups during collaborative scientific argumentation in order to improve the quality of a group outcome. It is also important to note that the type of group outcome desired (e.g., a better understanding of topic, a high quality argument, a better understanding of collaborative scientific argumentation, etc.) will, in all likelihood, affect the types of interactions that will need to be cultivated with a group as well. This indicates that more research that targets a specific type of group outcome and explores the underlying reasons for inter-group variation in performance during collaborative scientific argumentation is warranted and needed. The lack of consensus in the literature also suggests that this type of research, at least for now, should incorporate a method that will enable researchers to identify differences in the nature of the interactions that take place within high and low performing groups in a more holistic and emergent manner rather than imposing an existing analytical framework (such as Toulmin's Argument Pattern) a priori. This type of research, we argue, should help facilitate the identification and documentation of additional important aspects of collaborative scientific argumentation that might have been missed in other studies and will help clarify the field's understanding of the types of interactions that tend to lead to a desired group outcome during a specific type of instructional task. The findings from this type of research should also promote and support the development of new and more nuanced instruments or analytical frameworks that can be used by science educators to assess students' ability to participate in collaborative scientific argumentation and how students abilities to participate in this complex practice change over time.

In light of these issues, our goal for this study was to examine the types of interactions that seem to promote or constrain the development of a high quality written argument (one type of group outcome) during a task that was designed to promote and support collaborative scientific argumentation. To accomplish this goal, we decided to examine the ways individuals in two high and two low performing groups interacted with each other, ideas, and the available materials in order to identify and document potential differences that seem to be related to group performance. We decided to focus on four contrasting groups because it enabled us to examine the ways students engaged in collaborative scientific argumentation in greater detail than is generally feasible with a large sample. More specifically, focusing on four groups allowed us to apply an iterative analytical approach called verbal analysis (Chi 1997) to identify and then document substantial differences in the ways these groups engaged in argumentation as these differences emerged from the data and then connect them to the desired group outcome (i.e., a high quality written argument that includes a sufficient and accurate explanation and is well supported with appropriate evidence and reasoning). With this overview in mind, we now turn our attention to the theoretical framework and the empirical foundation that informed the design of this study.

Theoretical Framework

In order to guide our work, we have adopted a view of argumentation in science as a knowledge building and validating practice where individuals propose, support, critique, and refine ideas in an effort to understand 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 and Osborne 2002, p. 41) rather than as an effort to "justify or refute a particular standpoint" (van Eemeren et al. 2002, p. 38) or as the articulation of informal reasoning (e.g., Perkins et al. 1991; Sadler 2004; Zohar and Nemet 2002). We also differentiate between terms such as explanation, argument, and argumentation in our work. Explanations are statements that explicate or describe natural phenomenon. Arguments provide and support an explanation or other type of claim (such as the appropriateness of an experimental design, model, or research question) and can be either spoken or written. Argumentation is a process of establishing or validating a conclusion on the basis of reasons or the act of proposing, supporting, evaluating, and refining the process, context, or products of an inquiry (i.e., conclusions, explanations, or arguments). When conceptualized in this manner, scientific 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). The ways individuals or groups engage in scientific argumentation, and the nature of the products (i.e., explanation or arguments) that are produced through scientific argumentation, however, are strongly influenced by the goals, norms, and epistemological commitments of the scientific community because these goals, norms, and shared commitments shape what counts as quality for scientists.

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

This type of theoretical framework also stresses the importance of understanding how specific activities influence learning and the kinds of learning or knowing that are afforded and valued within the context or culture of a classroom. From this perspective, classroom activities, that are designed to engage students in specific types of practices (such as scientific argumentation) are not only important in terms of differences in their effectiveness at supporting the development of content knowledge, but also because the ability to engage in these practices and how to interact, communicate, and reason with others during these practices, is a fundamental part of what students learn (Scott et al. 2007). Therefore, if a goal of science education is to help students learn how to engage in scientific argumentation in a more constructive manner and to learn new content and modes of interacting or reasoning as part of the process; science educators must learn how to structure the activities that take place within the classroom so they are productive (i.e., result in a desired group outcome) and educative (i.e., help students learn more about the process of collaborative scientific argumentation). In order to accomplish this goal, however, more research is needed that identifies the types of activity structures or approaches that are conducive for learning about (i.e., processes) and from (i.e., content) scientific argumentation, the challenges students face when they engage in these types of activities (so teachers can help students negotiate these issues), and the types of interactions that take place during these activities that tend to result in the desired group outcome (so they can be cultivated by teachers).

Empirical Foundation and the Focus of the Current Study

Science educators have developed numerous ways to engage students in collaborative scientific argumentation over the last decade (see Clark and Sampson 2006b; deVries et al. 2002; Kelly et al. 1998; Kuhn and Reiser 2006; Osborne et al. 2004; Sadler et al. 2007; Sampson and Clark 2009; Sandoval and Reiser 2004 for examples). There are, however, a number of challenges associated with these types of activities that students must overcome in order to be successful that have been uncovered in this literature. These challenges can be organized around two broad themes. One theme reflects the various challenges that are associated with the nuances of scientific argumentation and the other theme stems from the additional challenges that students face when they engage in collaborative work.

Challenges Associated with Scientific Argumentation

Students encounter numerous challenges when they engage in argumentation activities that are associated with the nuances of scientific argumentation and the nature of scientific arguments. These challenges often stem from how the process and products diverge from the forms of argumentation they encounter in daily life rather than from a lack of skill or natural ability. For example, when students are asked to generate an explanation for why or how something happens, students must first *make sense of the phenomenon* they are studying based on the information available to them. Current research suggests that students struggle with this process (Abell et al. 2000; Kuhn and Reiser 2005; Sandoval 2003; Vellom and Anderson 1999) and often rely on their personal beliefs or past experiences to do so. Another challenge students face when engaged in scientific argumentation is the process of *generating a sufficient and useful explanation* that is consistent with the types of explanations valued in science (Carey et al. 1989; Lawson 2003; Ohlsson 1992; Sandoval 2003). Once students have generated a suitable explanation, students also have difficulty *justifying their explanation using appropriate evidence and reasoning* from a scientific perspective. Research indicates that students often do not use appropriate evidence, enough evidence, or attempt to justify their choice or use of evidence in the arguments they produce (Bell and Linn 2000; Erduran et al. 2004b; Jimenez-Aleixandre et al. 2000; Kuhn and Reiser 2005; Sadler 2004; Sandoval 2003). Finally, students often do not *evaluate the validity or acceptability of an explanation* for a given phenomenon in an appropriate manner. Current research indicates that students often do not use criteria that are consistent with the standards of the scientific community to determine which ideas to accept, reject, or modify (Hogan and Maglienti 2001; Kuhn and Reiser 2006) and distort, trivialize, or ignore evidence in an effort to reaffirm a misconception (Clark and Sampson 2006a; Kuhn 1989). Overall, this literature indicates that students often struggle with many aspects of scientific argumentation in spite of the skills they demonstrate when supporting or refuting a viewpoint in everyday contexts (Eisenberg and Garvey 1981; Schwarz and Glassner 2003; Stein and Bernas 1999; Stein and Miller 1991).

Challenges in Terms of Collaboration

To help address these challenges, many science educators (e.g., Abell et al. 2000; Bell and Linn 2000; Kuhn and Reiser 2005; McNeill et al. 2006; Schwarz and Glassner 2003) organize students into collaborative groups to promote and support better scientific argumentation. These authors suggest that opportunities to collaborate with others can lead to better products (e.g., written explanations or arguments and other artifacts) or more fruitful intellectual outcomes (e.g., a more sophisticated understanding of the topic, new ideas to pursue, or a deeper awareness of the contexts and processes of an investigation) because groups can pool knowledge and take advantage of different cognitive and monitoring resources. Encouraging students to collaborate with others, however, does not guarantee a high quality product or other useful intellectual outcome. Several descriptive studies suggest that the ways group members interact with each other, what they discuss, and how a group attempts to accomplish a task can profoundly influence group products and outcomes. These studies illustrate the additional challenges that some students face when asked to collaborate with each other during tasks that require student to propose, support, evaluate, and refine ideas.

The work of Richmond and Striley (1996), for example, indicates that individual group members can have a substantial (and often negative) impact on the products produced by a group. Richmond and Striley found that "alienating leaders" often have a negative influence on group explanations. Alienating leaders are individuals who (a) are confident that their explanation is correct, (b) are not interested in hearing what others have to say, and (c) impose their ideas on others. Richmond and Striley noted that alienating leaders "not only controlled the ways group members were able to participate in the work but also shaped the definition of the work to be done" (p. 852). Richmond and Striley found that the explanations generated in groups with alienating leaders were inferior and the process of building these (explanations) was just a procedural one.

In another case study, which focused on the level of reasoning expressed in conversation during small group discussions about the nature of matter, Hogan et al. (2000) highlighted the importance of discourse that focuses on collaborative knowledge-building. They describe collaborative knowledge building discourse as discussion that (a) focuses on "scientific phenomena and ideas" (p. 387), (b) follows one of three major patterns (consensual, responsive, or elaborative), and (c) is characterized by high rates of queries, the presentation of new ideas, and comments that indicate agreement. These kinds of responses, according to Hogan, Natasi, and Pressley, serve to prolong the discussion of ideas and lead to higher levels of reasoning. They noted that the more successful groups in their study were able to sustain knowledge-building dialogue on their own while the least successful groups were not. As in Richmond and Striley's study, students in groups that interacted with each other's ideas more and did not engage in a competition seemed to develop a better intellectual outcome (in this case, level of reasoning).

More recently, Southerland et al. (2005) highlighted the important role of rhetorical moves and persuasion in shaping meaning-making when groups attempt to generate an explanation based on data. Southerland and colleagues examined how two small groups of students in an urban classroom attempted to make sense of condensation as part of an inquiry unit on the water cycle. They found that evaluative comments about the contributions of other group members were extremely persuasive within these groups and influenced the nature of the collaborative work. These types of comments, apart from students' academic status, were shown to be an important influence in not only social knowledge production but also in individual internalization. Thus, the quality of a group product or other intellectual outcome generated by a group might be strongly influenced by the number of challenges and critiques that occur within a particular group. Other research, however, indicates that the emergence of this type of discourse requires students to be comfortable with this type of interaction pattern and each other (Carlsen 2007), which seems to be another challenge associated with collaboration.

The Focus of the Current Study

The literature reviewed here illustrates a number of challenges associated with scientific argumentation and collaborative work that can constrain the attainment of a desired group outcome. The specific nature and impact of the interactions that promote better group outcomes, however, remain less well understood. Overall, this literature clearly indicates that more research is needed that examines the interactions of social, cognitive, and contextual factors that can constrain, and perhaps more importantly, foster the development of a desired group outcome during collaborative scientific argumentation. This literature also indicates that more research that seeks to understand how these interactions may differ depending on the nature of the task and the targeted group outcome. The current study addresses these issues by examining group interactions and processes that seem to promote or constrain the development of a high quality written scientific argument during a collaborative scientific argumentation task that involves the evaluation of several alternative explanations for a discrepant event.

Method

This study examines the nature of the collaborative scientific argumentation that took place within four different groups of three students (i.e., triads). The four triads were chosen from a larger subset of 28 groups that took part in another study that examined the impact of collaboration during scientific argumentation on written argument quality and individual learning (Sampson and Clark 2009). The original study included an assessment of the written arguments produced by each group. That study used four criteria to assess the quality of a written argument: (a) the *sufficiency of the explanation* (i.e., the extent to which the explanation answers the research question), (b) the *conceptual quality of the explanation* (i.e., the accuracy of the various components of an explanation), (c) the *quality of the evidence* (i.e., the extent to which appropriate and relevant evidence was used to support the explanation), and (d) the *adequacy of the reasoning* (i.e., the extent to which a group linked the evidence to the explanation and justified the choice of evidence). We used these same criteria to define a high quality group product in the current study and to identify the two high and two low performing groups. More specifically, the desired group outcome of the collaborative scientific argumentation in this investigation was a high quality written argument that provided a sufficient and accurate explanation for a discrepant event and included appropriate evidence and adequate reasoning as support (as defined by the four criteria listed above). The full details of how we scored the written argument quality in this study are included in Sampson and Clark (2009).

Participants

The students in the original study attended a large suburban public high school located in the southwestern United States. These students ($N=168$) were drawn from six general chemistry classes taught by two teachers. Forty percent of the students in the original study were male and 60% were female. Approximately 71% of the students were White, 18% Latino/a, 4% 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 ($M = 15.77$, $SD = .62$).

The two most common instructional approaches used by the teachers in these classrooms were direct instruction and confirmatory laboratory activities. As a result, these students had a great deal of experience working in groups during the lab activities but little experience with collaborative problem solving, open or guided inquiry, and scientific argumentation. The decision to study a sample of students who did not engage in scientific argumentation, collaborative problem solving, or inquiry on a regular basis 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 work with a sample of science students who would be more representative of typical classrooms rather than a group of students who were well versed in collaboration and scientific argumentation. Thus, the results of this study should be viewed as a benchmark of what students can be expected to do when they are first asked to engage in these practices. The results also help illustrate the types of interactions that might lead to a high quality group outcome under these conditions.

The 28 groups in the original study were either all female or all male. Individuals were assigned to a same gender group in that study in order to help eliminate the impact of confounding variables in the quantitative analysis. The results of that study indicated a great deal of inter-group variation in performance but no difference in the overall performance between the male and female groups (Sampson and Clark 2009). With this information in mind, the four groups that are the focus of this study were selected using a two-step process. First, all 28 groups were rank ordered from high to low based on the quality of their written argument (12 points possible, *Maximum* = 11, *Minimum* = 3, *Mean* = 7.21, *Standard Deviation* = 1.64). Two groups (one female and one male) were then chosen from the higher and lower achieving groups as being representative of the interactions that took place in the other groups at those performance levels. As part of this process, we examined the videotapes to make sure that we were not studying an unusual case (e.g., a group that did not collaborate at all or the group went about their work in a unique manner). We also made sure that the selected groups included individuals with high and low levels of background knowledge. The four focal groups are described in the following paragraphs and in Table 1.

Group A was a high performing group that produced an argument ranked at the 89th percentile (9 out of 12 points) for group arguments. Their argument contained an accurate explanation that was well supported with appropriate evidence and reasoning. This group

Table 1 Characteristics of the group members in the two high and two low performing groups

Group	Characteristics of group members				
	Gender	Ethnic background	Home language	Percentile score on content knowledge pretest	Conversational turns
A (High performing)					
Student 1	M	White	English	14th	70 (36%)
Student 2	M	White	English	90th	78 (40%)
Student 3	M	White	English	35th	46 (24%)
B (High performing)					
Student 1	F	White	English	90th	60 (38%)
Student 2	F	White	English	46th	36 (23%)
Student 3	F	Latina	Spanish	7th	61 (39%)
C (Low performing)					
Student 1	F	White	English	85th	53 (42%)
Student 2	F	White	English	35th	55 (43%)
Student 3	F	Latina	Spanish	60th	29 (15%)
D (Low performing)					
Student 1	M	White	English	24th	33 (41%)
Student 2	M	Latino	Spanish	5th	27 (33%)
Student 3	M	White	English	90th	21 (26%)

consisted of three white male students that spoke English at home. While the group ranked at the 89th percentile in terms of their argument, these students' scores on the content knowledge pretest (*see* Procedure) ranged from the 14th to 90th percentile. Participation was relatively even. The student who contributed the least to the discussion made 24% of the turns, the middle student made 36% of the turns, and the individual who contributed the most made 40% of the turns (*see* Segmenting and Analyzing Transcripts).

Group B also produced an argument that scored at the 89th percentile (9 out of 12 points). This group consisted of two white females that spoke English at home and one Latina student who indicated that her home language was Spanish. The group members' content pretest scores ranged from the 7th to 90th percentile. The student who made the most contributions to the discussion made 39% of the turns, the middle student contributed 38% of the turns, and the individual who contributed the least made 23% of the turns.

Group C was a low performing group. Their argument ranked at the 21st percentile (5 out of 12 points). Their argument contained a sufficient but inaccurate explanation that was supported with inappropriate evidence and inadequate reasoning. This group, like group B, consisted of two White female students who spoke English at home and one Latina student who spoke Spanish in her home. Content pretest scores ranged from the 35th to 85th percentile. The student who contributed the least to the discussion made 15% of the turns, the middle student contributed 32% of the turns, and the individual who contributed the most made 43% of the turns.

Group D also produced an argument that was scored in the 21st percentile (5 out of 12 points). This group, like group C, produced an argument with a sufficient but inaccurate explanation that was supported by weak evidence and reasoning. This group consisted of two White males and one Latino male. Content knowledge scores in this group ranged from the 5th percentile to the 90th percentile. The student who contributed the least made 26% of the turns, the middle student made 33% of the comments, and the individual who contributed the most to the discussion made 41% of the turns.

All four groups were highly engaged with the problem and almost all comments were related, in one way or another, to the task. Membership in each group represented roughly similar spreads of content pretest scores. Each group also showed a roughly similar distribution of turn taking among its members. In addition, reviews of the videos suggested that the groups appeared to be safe social contexts for the individual group members. The students seemed to get along well and the interactions that took place within all four groups were friendly and respectful. Overall, the individuals in all four groups appeared to be comfortable interacting with each other and were highly engaged in the task.

The Task

All 28 groups in the original study completed a complex task designed to engage them in collaborative scientific argumentation. This task is called the *ice melting blocks problem.* This problem requires students to generate an argument that explains why an ice cube sitting on a block made of aluminum melts faster than an ice cube sitting on a block made of plastic. The students were supplied with the two blocks, ice cubes, and three handouts at the beginning of the task. The first handout included the introduction to the problem, a focus question ("Why does the ice melt faster on block A?"), and two prompting questions designed to motivate students to generate an argument that provided their explanation ("What is your explanation?") and a justification for that explanation ("How do you know? In this space below, defend your explanation with appropriate evidence and reasoning."). The second handout included a list of six different plausible explanations that the students could use in their argument. The students were told that they could use one of the supplied explanations, a modified version of one of the explanations, or a completely original explanation. The third handout included information about the blocks (i.e., initial temperature, mass, density, time for it to change temperature in different environments, etc.) and the molecular-kinetic theory of matter that the students could use to critique the explanations and to generate the evidence and reasoning necessary to justify one explanation over the others. The students were given approximately 40 min to discuss the various explanations and then generate an argument that provided and supported the most valid or acceptable explanation for their observations. All of the groups were able to complete the task in the allotted time.

Procedure

The intervention consisted of four 1-hour sessions that spanned four consecutive school days. In the first session, the participants completed a content knowledge pretest to characterize the background knowledge of the group members. The classroom teacher administered this assessment. This information was used in the original study (see Sampson and Clark 2009) to ensure the equivalence of the individuals assigned to each condition. This information was used in this study to control for the possibility that the individuals in the higher performing groups knew more about the topic than the individuals in the lower performing groups. All four focal groups, as noted earlier, included individuals with high levels and individuals with low levels of content knowledge as measured by the pretest.

The first author introduced the students to the nature of scientific arguments during the second session by taking on the role of the classroom teacher. To accomplish this task, we used a framework (see Fig. 1) that was inspired by the work of Toulmin (1958) and other

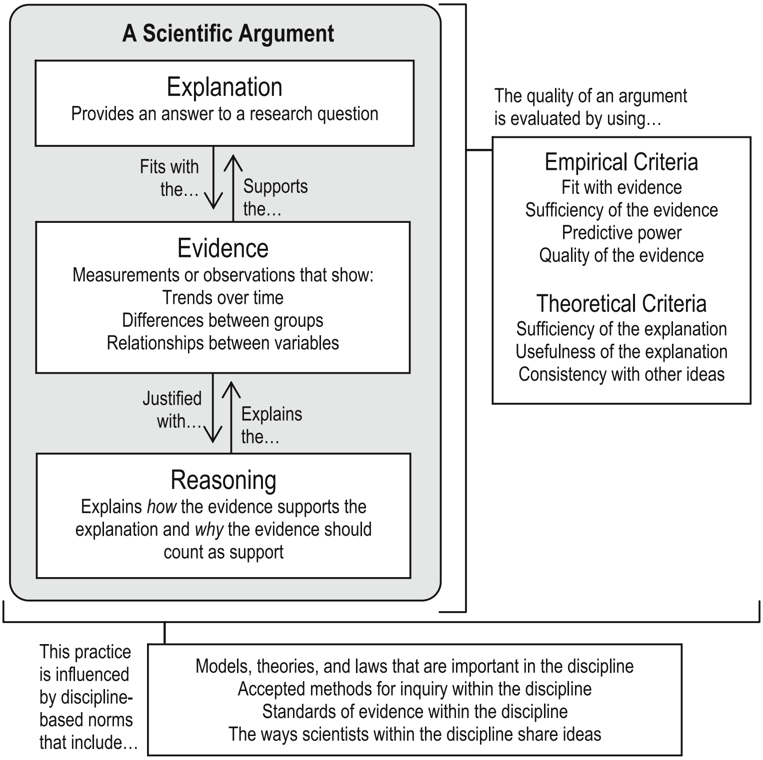

Fig. 1 The argument framework used in this study

researchers working to integrate scientific argumentation into the teaching and learning of science (Kuhn and Reiser 2005; McNeill et al. 2006; Sandoval and Millwood 2005; Stewart et al. 2005). This framework, which reflects our theoretical perspectives on argumentation, breaks down a scientific argument into three interrelated components: an explanation (similar to Toulmin's claim), evidence (similar to Toulmin's data), and reasoning (a combination of Toulmin's warrants and backings).

The *explanation* component of the framework proposes a potential answer or causal mechanism in response to the research question that guides the investigation. Depending on the question, 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 provides 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, in order for this information to be considered 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 provides a rationalization for 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 (2008) have adapted Toulmin's model in a similar manner to guide their curriculum development efforts. Their work suggests that this type of framework is an appropriate and productive way to introduce students to the nature of scientific arguments.

This framework also highlights some rigorous criteria that can be used to evaluate the quality of a scientific argument. Based on the work of Stewart and colleagues (Cartier and Stewart 2000; Passmore and Stewart 2002; Stewart et al. 2005), our framework calls for students to use both empirical criteria (e.g., how well the explanation fits with available data, the predictive power of the explanation, the sufficiency of the evidence cited, and the quality of evidence) and theoretical criteria (e.g., how consistent the explanation is with other scientific knowledge and whether or not the explanation is a useful way to think about the phenomenon) to assess a scientific argument. This component of the framework is designed to help students understand the types of criteria that are valued and used in science (Hogan 2000; Kuhn and Reiser 2006; Sandoval 2003) and to provide tools that students can use to evaluate scientific explanations in a more productive manner (e.g., Schwarz and Glassner 2003).

This framework was used during the second session to (a) provide students with a template that highlighted the different components of an argument (e.g., explanation, evidence, and reasoning), (b) make what counts as quality for each aspect of the template explicit for the students (e.g., quality evidence consists of observations or measurements that show difference between groups or objects), and (c) give the students an opportunity to develop a basic understanding of some of the criteria that scientists use to evaluate arguments (e.g., explanations should fit with the available data). These curricular goals were met by first defining each component of the framework and by making the rationale behind each component explicit. The first author then modeled how to construct a scientific argument and provided examples of both strong and weak arguments for the students to critique.

The students completed the *ice melting blocks* problem during the third session without input from the researchers. The triads worked at a table in front of a video camera so that the interactions that took place between the students and the available materials could be recorded. The triads completed the task in empty rooms during this session with an adult present to operate the video camera in order to ensure video and audio quality. The current study focuses on the data collected during this session of this intervention.

During the fourth and final session, all the participants completed a mastery version of the *ice-melting blocks* problem and a transfer problem on their own. The mastery version of the *ice-melting blocks* 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 an 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. This transfer problem, which is conceptually similar to the ice-melting blocks problem, required the students to generate a written argument that explains why metal and plastic objects feel like they are different temperatures although they have been sitting in the same room for a longer period of time. Students received a handout that included directions, the same prompting questions, and a data set. The students were given 40 min to complete these two problems. All of the participants completed both problems in the allotted time. The classroom teacher was once again responsible for the administration of these assessments.

Segmenting and Analyzing Transcripts

The conversations that took place within the triads were video recorded and transcribed. The transcription focused specifically on the sequence of turns and the nature of the interactions rather than speaker intonation or other discourse properties. Transcripts were parsed into turns, which were defined as segments of speaker-continuous speech. If an interruption stopped the speaker from speaking, the turn was considered complete, even if the content of the turn was resumed later in the conversation. If the student did not stop talking even though someone else was speaking, then all of the content was considered to be part of that same turn. One-word utterances, such as “yeah”, “uhm”, and so on, were also considered to be turns.

We relied on a two-stage analytical approach based on the verbal analysis methods developed by Chi (1997) to identify and document the substantial differences in the argumentation practices that took place within the two higher and two lower performing groups. To do this, the videos and transcripts of the conversations were first examined in a holistic manner in order to identify potential differences in the ways these groups engaged in collaborative scientific argumentation. The theoretical framework and the themes we outlined in our review of the literature guided our observations during this step of the process. The theoretical framework helped focus our attention on the ways these groups engaged in argumentation to solve problems and advance knowledge and to determine if the students’ argumentation practices were aligned with the goals, norms, and epistemological commitments of science (i.e., the culture of science). The themes we extracted from the collaboration and argumentation literature, on the other hand, helped focus our attention on the cognitive, contextual, and social aspects of the collaborative scientific argumentation and the ways these students’ practices were influenced by the nature of the task and tools available. Once a potential difference was identified though this more holistic step of the analytic approach, we then developed a coding scheme or adapted from the available literature to measure and document the extent of these differences. We retained and reported the differences that were supported by the data and discarded the prospective differences that were not well supported. This iterative and emergent process of identification and quantification of prospective differences continued until no more potential differences could be found in the videos or transcripts.

Results

We identified five substantial differences in the collaborative argumentation practices of the two higher and two lower performing groups that may help explain why the higher performing groups ended up creating stronger arguments (our desired outcome) than the lower performing groups. Some of these differences provide new insights into group mechanics in this context while others replicate findings from earlier studies. These differences included (1) the number of unique content-related ideas introduced into the discussion, (2) how group members responded to proposed ideas, (3) the nature and function of contributions during discussion episodes, (4) the criteria used to justify and evaluate explanations and ideas, and (5) how groups used the available corpus of data. Although we chose to present these differences sequentially, it is important to note that these differences appear interrelated (a point that we will return to in our discussion of these results). The specifics of the coding schemes that were used to document and measure the extent of these differences are described in conjunction with the analyses of the specific differences between groups. These descriptions also provide other relevant information about the analyses such as the inter-rater reliability of each coding scheme.

Difference 1: The number of unique content-related ideas introduced into the discussion

The first notable difference between the groups involved the number of unique content-related ideas introduced during the discussions. Individual group members proposed some of these ideas in an unprompted manner (e.g., "I think the ice melted faster on 'A' because it is metal") or as they read the alternative explanations supplied to them whereas other ideas seemed to emerge through the interactions that took place between the group members (e.g., "but how could that be, metal is a good insulator"). Regardless of the source, however, these ideas were important to the conversation because they provided possible solutions to the problem that could be accepted, rejected, or modified by the group members. The individuals in the higher performing groups voiced a greater number of ideas than the individuals in the lower performing groups.

To document this trend, a coding scheme was developed to identify the unique content-related ideas that were spoken aloud during the discussion. This coding scheme relied on the same facet analysis approach (Hunt and Minstrell 1994; Minstrell 2000) that was used to score the conceptual quality of the written arguments in the original study (see Sampson and Clark 2009). Facets are ideas that lack the structure of a full explanation and can consist of nominal and committed facts, intuitive conceptions, narratives, p-prims, or mental models based on experiences at various stages of development and sophistication (Clark 2006). Examples of content-related ideas that were identified in this analysis include inaccurate facets of student thinking such as 'block A was initially colder than block B', 'metal absorbs cold', and 'metal is a good insulator' and accurate facets such as 'heat flows from hot to cold objects' and 'metal is a good conductor'. Two coders independently coded all four discussions. The inter-rater reliability (i.e., percent agreement between the two coders) for the identification of unique ideas introduced to the discussion was 90% (Cohen's Kappa = .81).

The average number of unique content-related ideas introduced into the conversation by the higher and lower performing groups was then compared. As shown in Fig. 2, the higher performing groups voiced twice as many unique content-related ideas (30.5) as the less productive groups (15.0). The individuals in these groups not only mentioned more ideas (e.g., metal absorbs cold) from the materials supplied to them (i.e., the list of explanations and the data sheet) as they worked, they also introduced more ideas from personal experiences or other contexts (e.g., They block heat... like a cooler). This, as highlighted earlier, is a potentially important difference. When students engage in collaborative scientific argumentation, they have an opportunity to exchange ideas, challenge each other, and discuss or

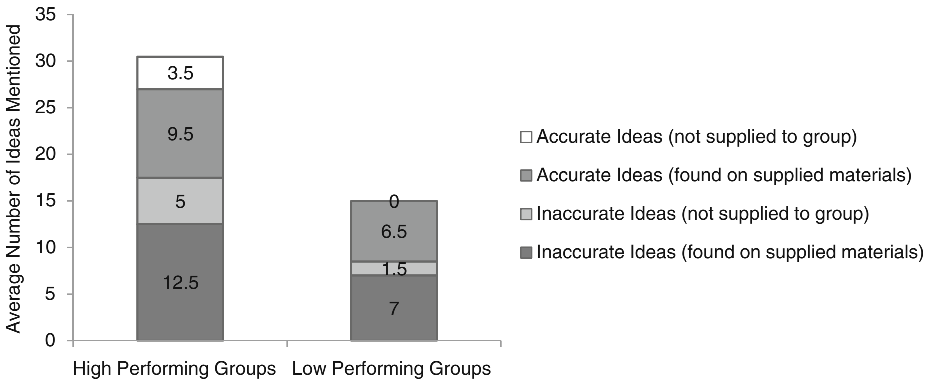

Fig. 2 Average number of unique content related ideas voiced by group members working in the high and low performing groups

explore possible solutions. If few ideas are introduced into a discussion, however, there is less to discuss and the potential benefits of exposure seem to be diminished.

Difference 2: How group members responded to proposed ideas

The second key difference in the collaborative interactions of the groups involved the ways in which group members responded to the ideas introduced during the discussion. Students in the higher performing groups were more likely to discuss an idea before it was accepted or rejected while the students in lower performing groups were more prone to accept or reject ideas without discussion. This difference suggests that the individuals in the higher performing groups were attending to the thoughts of the other group members more often than the individuals in the lower performing groups. This is an important difference because the literature that is devoted to understanding collaborative group work suggests that group performance is often hindered during tasks that require the generation of a group product when group members do not discuss each other's ideas (e.g., Barron 2000, 2003; Hatano and Inagaki 1991; Hogan et al. 2000; Kuhn and Reiser 2005; Rochelle 1992).

To document this trend, a coding scheme was used to record how students responded to the ideas proposed during a discussion based on the work of Barron (2000, 2003). Four categories of responses were used: *accept*, *discuss*, *reject*, and *ignore*. *Accept* responses included any reaction where an individual voiced agreement with the speaker, supported the proposal, or incorporated the idea into the group product. *Discuss* responses included any reaction that resulted in further discussion of the idea. Examples of this type of response include questioning the rationale behind an idea, challenging it with new information or a different idea, asking for clarification, and revising or adding to an idea. *Reject* responses included any reaction that voiced disagreement with the speaker or made a claim that an idea was incorrect. Finally, *ignore* responses were coded as not giving a verbal response to an idea when it was proposed. Definitions and examples for each code are provided in Table 2. Two coders independently coded all four discussions. The inter-rater reliability for the identification of the type of reaction to a proposed idea was 87% (Cohen's Kappa = .80).

Because the number of ideas proposed differed in each group, the results of this analysis are presented as percentages. Figure 3 shows the percentage of reaction types to the

Table 2 Codes used to examine the ways group members respond to proposed ideas

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

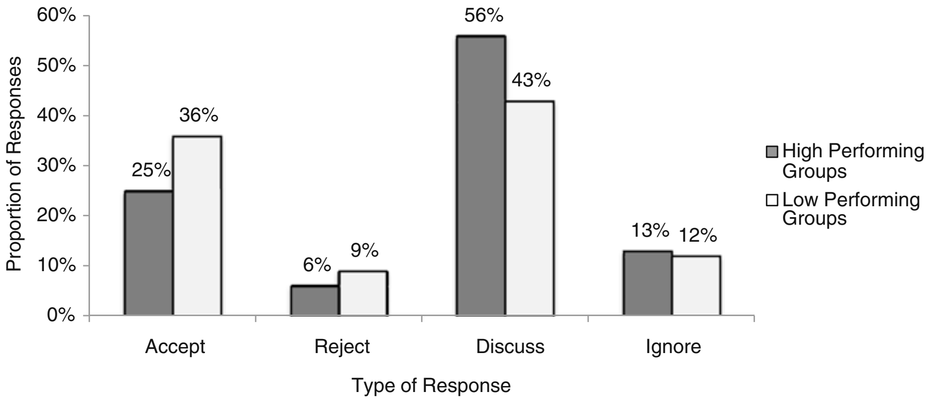

Fig. 3 Differences in the ways group members in high and low performing groups responded to the ideas proposed by other group members

proposed ideas in the higher and lower performing groups. As this figure shows, the higher and lower performing groups had about an equal percentage of *ignore* responses. However, the two higher performing groups were more likely to *discuss* ideas when they were proposed (56% of the time) than the two lower performing groups (43% of the time). On the other hand, the lower performing groups were more likely to *accept* or *reject* an idea without discussion (45% of the time) than the higher performing groups (31% of the time). The lower performing groups *accepted* proposals without discussion 36% of the time and *rejected* proposals outright 9% of the time whereas the higher performing groups only *accepted* a proposal without discussing them first 25% of the time and *rejected* them outright 6% of the time.

This data indicates that there was a notable difference in the ways individuals in the higher and lower performing groups responded to ideas when they were introduced into the conversation. The higher performing groups seemed to treat ideas as objects of cognition (Kuhn 1993; Mason 2001) that needed to be questioned, evaluated or revised more often that the individuals in the lower performing groups. Ideas in the lower performing groups, on the other hand, were accepted or rejected outright more often than they were in the higher performing groups. This may have prevented the conversation from moving forward or from turning into a more in-depth discussion of the underlying content. In other words, ideas in the two higher performing groups were often treated as a starting point rather than as an end point. To illustrate this difference, consider the following examples.

In the first example, taken from one of the lower performing groups, a number of ideas are proposed but these ideas are either accepted or rejected without discussion.

Student C-3: I think it's the second one because it says that Block A being placed on the table was—

Student C-1: Yeah, 'cause it's not [explanation] number four.

Student C-2: Okay. [reads explanation #2] The ice melts faster on Block A because Block A is a good conductor of heat. Although both blocks are the same temperature, heat energy transfers into the ice... That's not it.

Student C-1: I think its number three.

Student C-2: Yeah, me too.

Student C-1: [reads explanation #3] The ice melts faster on Block A because Block A is a good conductor. Although Block A is colder than Block B, it is still warmer than the ice. As cold moves into Block A, the ice warms up and melts. The ice on Block A melts faster because the cold moves from the ice into the block faster.

Student C-2: Yeah, that's it.

These types of "yeah, that's it" accept responses or "that's not it" reject responses were common in the dialogue that took place within the lower performing groups. As a result, these students rarely engaged with the content of an explanation in a substantial way or examined the underlying reasons for or against a particular viewpoint; instead these groups seemed to spend a majority of their time indicating that they were either for or against a particular explanation.

In contrast, when an idea was proposed in the higher performing groups, it often served as a starting point for a more in-depth discussion as illustrated in the following example.

Student B-2: [reading explanation #6] Plastic is a better insulator than metal, which is a good conductor. This means that Block B, which is made of plastic, retains all its energy. While Block A, which is a—made up of metal, transfers all its heat into the ice.

Student B-1: Do insulators block or retain heat?

Student B-3: They block heat... like a cooler.

Student B-1: 'Cause that's the—

Student B-2: 'Cause this one obviously melted a lot faster [points to block A]. This one's still [points to the ice on block B], like a cube.

Student B-1: Ice is a conductor or, I mean, metal's a conductor, right?

Student B-3: Mm-hmm.

Student B-1: Okay. So, so aluminum and then—so, let's see. If this one's—

Student B-3: This one's an insulator [points to block B], so it's—this is gonna take a really long time for it to melt.

Unlike the previous example, these students did not accept or reject the explanation outright. Instead the response of student B-1 led to a more in-depth discussion of the core concept involved, thermal conductivity. Elaborative responses of this type to proposed ideas appear to contribute to productive collaborative scientific argumentation because responses to proposed ideas that stimulate more conversation provides more opportunities for groups to evaluate, revise, or modify ideas. When groups reject or accept ideas without discussing them first, some of the potential benefits of engaging in scientific argumentation with others may be lost.

Difference 3: The nature and function of contributions during the discussion episodes

The third notable difference involved the overall nature of the conversation during the discussion of the merits of an idea in the *discuss* episodes. Individuals in the higher performing groups contributed a greater proportion of comments that challenged each other's ideas and a smaller proportion of comments devoted to seeking information, proposing and justifying their own ideas, or supporting the ideas of others when compared to the comments contributed by the individuals in less productive groups during these *discuss* episodes. This is an interesting difference because, as noted earlier, some authors (e.g., Andriessen et al. 2003b; Boulter and Gilbert 1995; Veerman 2003) have suggested that oppositional discourse can constrain the meaning-making process while others (e.g., Kuhn and Udell 2003; Osborne et al. 2004) have suggested that oppositional discourse is an important aspect of collaborative scientific argumentation. It seems that a greater proportion of oppositional comments during these *discuss* episodes may have stimulated more in-depth discussion of the idea involved and contributed to the generation of a better group argument in this case.

In order to confirm this potential difference, a coding scheme was developed to capture the overall nature and function of the contributions to the discussion during *discuss* episodes. Contributions can involve *task-related* or *off-task* comments. Very few *off-task* comments were observed in any of the groups during these episodes. *Off-task* comments are therefore not included in the analysis. The *task-related* comments were coded using four different categories: *information-seeking*, *exposition*, *oppositional*, and *co-construction*. *Information-seeking* comments request more information or clarification from others. *Exposition* comments propose, clarify, or justify ideas. *Oppositional* comments challenge the ideas of others. These include simple disagreements and disagreements accompanied by critiques. Finally, comments coded as *co-construction* summarize, revise, support, or add to the ideas of others. The unit of analysis for these codes was conversational turns within a *discuss* episode. The start and end point of a *discuss* episode was defined as the first

Table 3 Codes used to examine the overall nature and function of the contributions during *discuss* episodes

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

comment after an introduced idea and the first comment that indicated a new topic of discussion. Table 3 provides more detail about this coding approach. Once again, two coders independently coded all four discussions. The inter-rater reliability of this coding scheme was 86% (Cohen's Kappa = .80).

The results of this analysis are once again presented as percentages. Figure 4 shows the percentage of comments contributed by individuals in the two higher and two lower performing groups devoted to *information-seeking*, *exposition*, *opposition*, and *co-construction* during *discuss* episodes. As shown, the high and low performing groups had very similar proportions of the conversation devoted to *exposition* (24% and 21%, respectively). However, the individuals in the higher performing groups contributed a greater proportion of *oppositional* comments to the discussion (22% of the comments) than the individuals in the lower performing groups (10% of the comments). The individuals in the lower performing groups, on the other hand, contributed a greater proportion of comments that were coded as *information-seeking* (22% compared to 17%) and co-construction (47% compared to 37%) than the individuals in the higher performing groups. Overall, this data suggests that individuals in the higher performing groups contributed a greater proportion of oppositional comments when discussing the merits of idea during collaborative scientific argumentation. These disagreements and critiques appear to have influenced the quality of the argument produced by these groups.

To illustrate this trend, consider the following example episodes. The first example is representative of the overall nature of the discourse that took place between individuals within the lower performing groups when discussing the merits of an idea. The conversation in this example focuses primarily on the co-construction of knowledge. In other words, during this episode these students are revising, supporting, and adding to each other's ideas as they work to reach consensus.

Student C-1: I think its [explanation number] three 'cause metal—this is a better insulator, right? [Points to block A, which is made of aluminum]

Student C-2: Than this? [Points to block B, which is made of plastic]

Student C-1: Yeah.

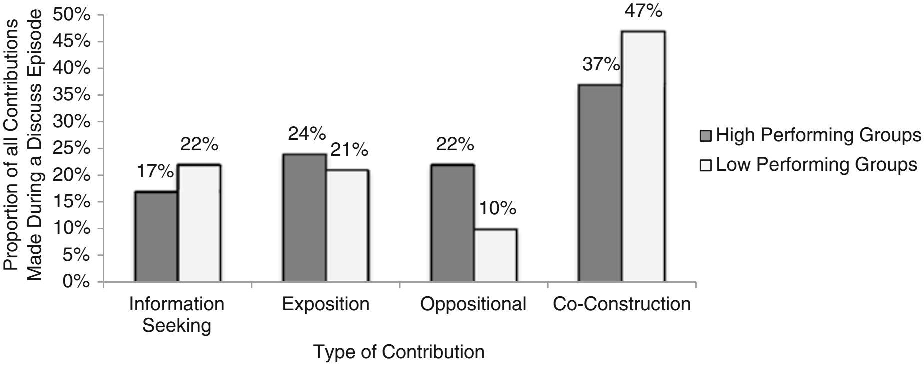

Fig. 4 Differences in the types of task-related comments contributed by individuals in high and low performing groups during *discuss* episodes

Student C-2: Right.

Student C-3: Wouldn't it be kind of like your thermos?

Student C-1: Oh, yeah, yeah. Like with heat—

Student C-3: Right, 'cause you wouldn't use a plastic thermos.

Student C-2: Yeah.

Student C-1: Yeah, that's true.

Student C-2: All right. So, it's not the forth one.

Student C-1: Yeah, 'cause it's not number four because it says, the ice melts faster on Block A because Block B is an insulator and Block B is not an insulator.

Student C-3: Yeah.

This excerpt is representative of the overall nature of the discussion that took place within the lower performing groups when group members did not accept or reject an idea outright. Although the task required the group members to evaluate several alternative explanations and *reject* responses directed at these explanations were common in these groups (e.g., "that's not it"), there were few instances where individuals in the lower performing groups actually challenged each other when discussing an idea. Instead, the students in these groups spent the vast majority of their time either elaborating on an idea and asking questions or agreeing with and supporting the ideas of the other group members. For example, in the episode above, rather than attempting to challenge the accuracy of an erroneous idea proposed by student C-1 (metal is a better insulator than plastic) or requiring student C-1 to support the merits of this idea, student C-2 simply asked a clarifying question ("than this?") and then voiced agreement ("right"). This was followed by a series of supporting comments offered by student C-3 ("Wouldn't it be kind of like your thermos", "...cause you wouldn't use a plastic thermos"). This type of interaction was common in the lower performing groups. These individuals seemed unwilling to disagree, challenge, or critique the ideas of other group members (even when an idea that was introduced into the discussion was inaccurate from a scientific perspective).

Now compare the above example with the following excerpt of dialogue taken from one of the higher performing groups. In this example episode, the discourse is more oppositional in nature.

Student A-1: It says, why—well, the question's, "Why does it—the ice melt faster?"

Student A-2: I think it's because of the temperature.

Student A-1: But it's colder.

Student A-2: And because of the—how it—

Student A-3: But if it's colder, it's not gonna melt.

Student A-1: It's not gonna melt faster.

Student A-3: I think it's 'cause metal conducts or takes in heat.

Student A-2: Mm-hmm.

Student A-1: Yeah. It absorbs heat.

Student A-2: And if it absorbs heat faster than it will melt faster?

Student A-1: Right.

Student A-3: Yeah.

This excerpt is representative of many of the exchanges that took place within the higher performing groups during *discuss* episodes. Students in these groups seemed much more willing to disagree, challenge, or critique the suggestions of other group members. Furthermore, oppositional discourse in these groups did not lead to the polarization of viewpoints or cause group members to opt out of the discussion. Instead, this type of discourse appeared to play an important role in moving the discussion forward. Disagreement and critiques often led to a critical examination of explanations, evidence, and reasoning. As a result, these types of interactions appeared conducive to the development of a stronger group argument.

Difference 4: The criteria used to justify and evaluate explanations and ideas

Another key difference involved the types of criteria that the group members used to judge the validity of explanations and ideas. The students in the higher performing groups frequently relied on rigorous criteria that were more appropriate for justifying and evaluating ideas in science, such as fit with evidence or consistency with other theories or laws. The students in the lower performing groups, on the other hand, typically used more informal criteria that were less appropriate for judging the quality of ideas and explanations in science, such as consistency with personal beliefs or personal experiences.

To document this trend, a coding scheme was developed to capture the types of criteria students used to justify or evaluate explanations and ideas. Two categories of criteria were used: *rigorous* and *informal*. *Rigorous* criteria include the reasons or standards that reflect the evaluative component of the argumentation framework outlined in Fig. 1. Examples of rigorous criteria include fit with data (e.g., “but the blocks are the same temperature”), sufficiency of data (e.g., “you do not have any data to support that”), coherence of an explanation (e.g., “how can something block and transfer heat at the same time”), sufficiency of an explanation (e.g., “that does not answer the question”), and consistency with scientific theories or laws (e.g., “according to the molecular-kinetic theory of matter, molecules are always in motion”). *Informal* criteria include less powerful reasons or standards for judging the validity of an idea in science. Examples of informal criteria include appeals to authority (e.g., “well that’s what she said”), discrediting the speaker (e.g., “he never knows what to do”), plausibility (e.g., “that makes sense to me”), appeals to analogies (e.g., “this is just like...”), fits with personal experience (e.g., “that happened to me once”), judgments about the importance of an idea (e.g., “that doesn’t matter”), and consistency with personal inferences (e.g., “it feels hotter so it is hotter”). Two coders independently coded all four discussions. The inter-rater reliability for the identification of the type of criteria used to justify or evaluate an explanation or idea was 93% (Cohen’s Kappa = .87).

Figure 5 shows how often individuals in the groups used rigorous and informal criteria (as a percentage of the total number of instances) to distinguish between explanations and to justify or evaluate ideas. As this figure shows, the higher performing groups relied on *rigorous* criteria 61% of the time and *informal* criteria 39% of the time. The individuals in the lower performing groups, in contrast, used *rigorous* criteria 31% of the time and *informal* criteria 69% of the time. The higher performing groups also used rigorous criteria to justify an idea (20% compared to 12%) and to challenge ideas (41% compared to 19%) more frequently than the lower performing groups. Overall, these data indicates that the

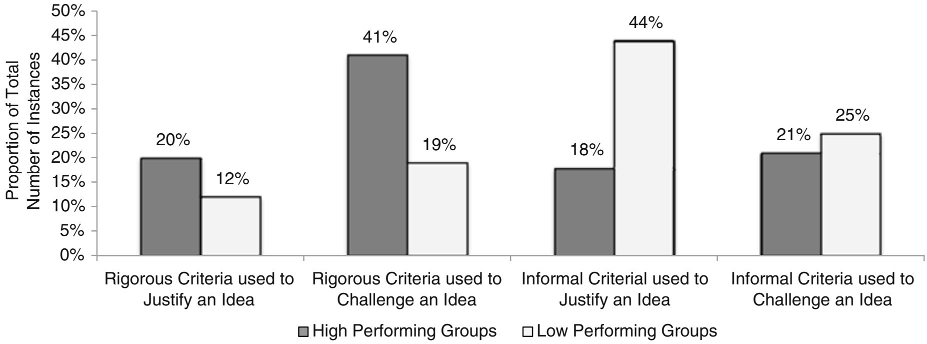

Fig. 5 Differences in the types of criteria used in high and low performing groups in order to challenge or justify ideas

individuals in two higher performing groups more likely to use *rigorous* criteria to support or evaluate ideas than the individuals in the two lower performing groups.

To illustrate this trend, consider the following examples. In the first example, students in a lower performing group rely on plausibility and personal inferences to evaluate the accuracy of an explanation. In other words, these students employed their personal viewpoints as the primary criterion for determining the validity or acceptability of explanations.

Student C-1: Okay. [Reads explanation #4] The ice melts faster on Block A because Block B is an insulator. As a result, energy does not move from the ice block—from the ice into Block B. So, the ice in Block B only melts because the air in the room warms it up. The ice melts on Block A because energy can move from the ice into Block A.

Student C-3: That sounds pretty good.

Student C-2: Yeah, that sounds good.

Student C-1: How come?

Student C-2: Because it's metal and metal absorbs cold.

Student C-1: Yeah. That makes sense.

These "that sounds good" or "that makes sense" comments were common in the dialogue that took place within the lower performing groups. The high frequencies of these types of comments suggest that these students were relying on informal criteria to distinguish between competing explanations or ideas. This potentially contributed to these groups' choices of explanations involving common misconceptions, such as "metal absorbs cold" or "insulators do not transfer heat."

On the other hand, as shown in the following example, the individuals in the higher performing groups relied on more rigorous criteria, such as fit with data, more often than the individuals in the lower performing groups when they were supporting and critiquing ideas.

Student A-1: So the explanation says that, it [the ice on block A] melts faster—'cause it's [block A] a conductor of heat.

Student A-3: Yeah, metal is a conductor of heat.

Student A-2: Right—because first it was the normal temperature, then—or it was—how do you say it? It was, like—first, it wasn't—it got colder. Meaning that the block had to be—

Student A-1: But just 'cause it got colder doesn't prove that it conducts heat.

Student A-2: Well, no, 'cause the—at first it was—[touches block A] it's warm now. Yeah, this one's warm now.

Student A-3: [touches block A] Yeah. Compared to what it was.

Student A-2: Is it—no, I mean, I think it is a conductor 'cause it can—[points to the data sheet] when it's put in the freezer, it got colder faster than the wood or plastic did. But then when it was put in the oven it also got hotter faster.

Student A-1: Oh... So it got hotter faster?

Student A-2: Mm-hmm.

Student A-3: Right, see it shows it right here [points to the data sheet].

Student A-1: Okay. So it [block A] is a better conductor of heat.

This excerpt is representative of many of the exchanges that took place within the higher performing groups. Students in these groups seemed to rely on more rigorous criteria to distinguish between competing explanations and ideas as they collaborated. Although students in the higher performing groups used personal viewpoints as criteria some of the time, the individuals in these groups more frequently employed criteria that aligned more closely with criteria valued in science. Potentially as a result, the two higher performing groups were more successful in identifying inaccurate ideas that seemed plausible but did not fit with the available data.

Difference 5: How groups used the available corpus of data

The final notable difference involved how the groups used the materials available to them. All four groups followed the same basic series of steps to complete the task. Each group first discussed which of the six alternative explanations was the most acceptable or valid. Then once the individuals in the group had reached a consensus, they worked together to craft a written an argument to defend the validity of the chosen explanation. The higher and lower performing groups, however, differed in the way they used the corpus of data that was provided to them during this process. The students in the lower performing groups only used the corpus of data to craft their written argument while the individuals in the higher performing groups used this information to help them evaluate the alternative explanations and to generate their written argument. This observation suggests that the types of materials made available and the ways that group members choose to use these materials when engaged in scientific argumentation may have a significant impact on the quality of the argument groups produce.

To document this trend, the video records of all four groups were examined in order to identify two milestones in the activity of the groups. One milestone involved the group's shift from evaluating the alternative explanations to generating their argument. The other milestone was the point when the group first used the data supplied them. Two coders independently identified these events in all four discussions. The inter-rater reliability for the identification of these two events was 100% (Cohen's Kappa = 1.00).

Figure 6 shows the amount of time each group spent evaluating the alternative explanations and generating their written argument as a percentage of the total. As

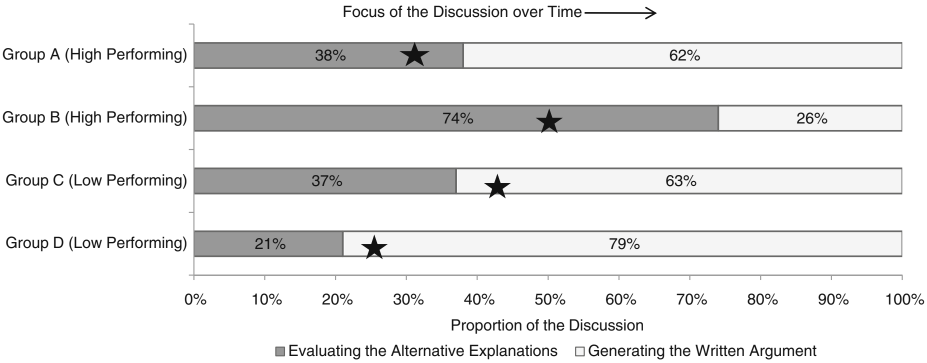

Fig. 6 Differences in the proportion of the time higher and lower performing groups spent evaluating explanations and generating their argument. The stars indicate the point in the conversation when a group member first used the available corpus of data to help accomplish the task

illustrated in Fig. 6, there was no major difference between the higher and lower performing group in terms of the percentage of time devoted to these two activities. In fact, group A (a high performing group) and group C (a low performing group) devoted approximately the same percentage of time to each step. There was, however, a major difference between the higher and lower groups in terms of when the groups chose to use the available corpus of data during the task. This is illustrated in Fig. 6 with stars (★). The stars indicate where in the discussion a group member first used the available data sheet. As can be seen in Fig. 6, individuals in the higher performing groups first began to use the data while they were still in the process of evaluating the alternative explanations whereas the individuals in the lower performing groups did not. This suggests that the individuals in the higher and lower performing groups were using the available data for different purposes during the collaborative argumentation process. The individuals in the higher performing groups used the corpus of data to evaluate the merits of the alternative explanations and to generate their arguments whereas the less productive groups only used the data to support a chosen explanation.

To illustrate this trend, consider the following example from a lower performing group. In this excerpt, the students reach consensus as to which explanation to use and then turn their attention to the generation of their written argument. As they do this, two of the group members begin to use the data sheets supplied to them as a way to find the evidence they need to defend their explanation.

Student D-2: I think it is number 3 or number 4

Student D-1: Yeah. I like this one.

Student D-3: Which one?

Student D-1: This one [points to #4]

Student D-3: And that block is aluminum [points to block A] and that one's plastic [points to block B]?

Student D-1: Yeah.

Student D-3: Okay.

Student D-1: All right. Who wants to write?

Student D-2: I will. Just look at the data tables while I write this down.

The individuals in the two lower performing groups apparently did not recognize the value of the data for evaluating the acceptability or the validity of an explanation. Instead, these individuals only used the materials as a way to justify their explanation. In other words, the materials were used as a way to defend their decision rather than as a way to make their decision.

The individuals in the higher performing groups, however, used the data to distinguish between explanations and to confirm the accuracy of an explanation *before* making their final decision. This trend is well illustrated by the following example. In this example, the group members use the data supplied to them as a way to evaluate one of the alternative explanations.

Student B-3: I think it is the last one. Do we all agree?

Student B-2: Yeah.

Student B-1: Yeah.

Student B-3: Okay. So now we need to find evidence to make sure it is right. [reads the data sheet] So block A is aluminum.

Student B-2: [Reads from the data sheet] Ten minutes, 14. [stops] So then, we should put that Block A, the ice melts faster because it's aluminum and—

Student B-3: No, don't do that [points to female #1 who was starting to write down an explanation].

Student B-1: Why not?

Student B-3: Well, we need to find evidence to make sure, like, this is true. Then we can go through and check it off.

Following this excerpt, the group members in this example spent several minutes looking over the data sheet and discussing various pieces of information. Their comments tended to focus on interpretations of the data and implications of data for their task. When they decided that something was relevant to their work, student B-3 made a mark next to that piece of information on her data sheet. Finally, after systematically going through the entire corpus of data, the following dialogue took place.

Student B-1: Now, can we write it [the explanation] down?

Student B-3: Well, we have to make sure everything matches up good and then do that. So, okay, for this it says that plastic is a better insulator than metal, which is a conductor. So, the plastic would retain the energy, which means it wouldn't melt the ice.

Student B-2: Mm-hmm.

Student B-3: So, that goes with...

Student B-2: So, like, the energy wouldn't transfer.

Student B-3: [checks off two pieces of data on the data sheet]...those.

Student B-2: Yeah. Ok, we're good. And then what's the next one.

Student B-3: Nearly identical temperatures between Block A and the room, but it still melts it faster. So, that means that metal would have to... [checks off another piece of data on the data sheet]

Student B-2: Yeah.

Student B-3: Plastic is an insulator. [checks off another piece of data on the data sheet]

Student B-1: Wait that doesn't make sense

Student B-3: That's why it took longer

Student B-1: Oh.

Student B-3: Those last two go with that. All right, now we can write it [the explanation] down.

These examples clearly illustrate how the supplied data were used differently in the two higher and two lower performing groups. In the lower performing groups the materials were only used to generate their argument whereas the higher performing groups used the supplied data throughout the process. This difference seemed to influence the ultimate quality of the groups' final written argument.

It is important to note here, however, that none of the groups returned to the process of evaluating the alternative explanations once they began the process of crafting their written argument. In these groups, once an explanation was agreed upon, a different explanation was no longer an option. This trend did not seem to have a negative impact on the quality of the written argument produced by the individuals in the higher performing groups because they had already eliminated the explanations that did not fit with the available data and knew which data to use to support their chosen explanation. The lower performing groups, on the other hand, often struggled to find data that supported their ideas and often came across counterevidence that contradicted an element of their chosen explanation as they attempted to generate their written argument. Yet, when this happened the individuals in the lower performing groups never returned to the list of alternative explanations to see if another one fit with the available data better; instead these groups either ignored the anomalous data or attempted to explain it away. The two lower performing groups, as a result, produced written arguments that included an inaccurate explanation that was supported by inadequate evidence and reasoning.

Discussion of the Results in Light of the Existing Literature

This section presents an overview and a discussion of the interactions that took place between individuals, ideas, and materials described in the previous section in light of the extant literature. Two of the differences provide new insights into the functioning of high and low performing groups in this type of context and three of the differences replicate and support findings from earlier studies.

The first of the differences involves the number of ideas introduced into the discussions. This is a relatively new finding and is a contribution of this study. The individuals in the higher performing groups voiced twice as many unique content-related ideas in their discussions. As a result, individuals in the higher performing groups had an opportunity to consider the merits of a wider variety of viewpoints as a group. The members of the lower performing groups, on the other hand, seemed to examine the various explanations on their own (rather than as a group) and did not voice as many different ideas to the group. As a result, the individuals in these groups had fewer opportunities to exchange ideas, challenge each other, and discuss or explore possible solutions in a collaborative manner. The chance to critique or evaluate alternative explanations or ideas for a given phenomenon with others, as suggested by Driver et al. (2000) and Bell and Linn (2000), seems to be an important component of collaborative scientific argumentation that results in the generation of a high quality written argument.

The second key difference, that groups members respond differently to contributions in high and low performing groups, is well known in the general collaboration literature (see Webb and Palincsar 1996 for a review) but still warrants discussion in this context. The results from this study indicate that individuals in the higher performing groups were more likely to discuss an idea before accepting, rejecting, or modifying than the individuals in lower performing groups as they worked to evaluate several alternative explanations and to generate an argument. As a result, the individuals in the two higher performing groups had additional opportunities to examine the merits of an idea and the reasons for either accepting or rejecting an idea than the individuals in the two lower performing groups. It seems that when group members accept or reject an idea without discussion during collaborative scientific argumentation, as was the case in the lower performing groups, opportunities to critique, clarify, add to, or revise ideas are limited. This finding also supports the findings of Barron (2000, 2003) who suggests that partner responsiveness is an important component of productive collaborative interactions during problem-solving tasks that are complex in nature.

The third key difference involves the overall nature of the conversation during discussion episodes that focused on the acceptability or validity of an idea. Individuals in the higher performing groups contributed a greater proportion of oppositional comments than individuals in the lower performing groups during these episodes. These oppositional comments, which include simple disagreements, critiques, and requests for justification, indicate that the individuals in the two higher performing groups were more willing to challenge the merits of an idea during the conversation. Although this is not a new finding, it does suggest that oppositional comments seem to play an important role during collaborative scientific argumentation. As suggested by a number of authors (e.g., Jimenez-Aleixandre et al. 2000; Kuhn and Udell 2003; Osborne et al. 2004), this type of interaction appears to stimulate a more in-depth conversation or the critical analysis of the merits of an idea. Overall, the results of this analysis indicate that this type of interaction seems to have a positive impact on group outcomes (in this case the development of a written argument that includes a sufficient and accurate explanation that is well supported by appropriate evidence and reasoning). This observation, however, might only apply to tasks where students are encouraged to evaluate several existing alternatives. Interactions that are devoted to exposition (i.e. proposing, supporting or clarifying one's own ideas) or to the co-construction of knowledge (i.e., summarizing, revising, supporting, or adding to the ideas of others) might be more valuable when students are asked to generate and then justify a completely original explanation that explicates or describes a natural phenomenon.

The fourth difference involves the types of criteria that the groups used to distinguish between ideas. The higher performing groups used rigorous criteria that were more appropriate for scientific inquiry, such as fit with data or consistency with other theories and laws, more often than the lower performing groups. As a result, these groups seemed to be able to identify and rule out explanations that contained common misconceptions more effectively than the lower performing groups (who often relied on more informal criteria such as plausibility or fit with personal beliefs). This difference, as suggested by authors such as Kuhn and Reiser (2006) and Linn and Eylon (2006), indicates that group performance during tasks that require the evaluation of alternative explanations may hinge on the types of criteria that group members use to judge the acceptability or the validity of these explanations. Although this finding is not new, it helps to confirm the work of other researchers that stress the importance of helping students learn to use more rigorous criteria to evaluate or critique knowledge claims in science.

The final difference involves how groups use the available data during collaborative scientific argumentation. The individuals in the higher performing groups used the available data to critique the alternative explanations and to generate their arguments while the individuals in the lower performing groups only used the supplied data to justify their explanations. In other words, the individuals in the lower performing groups used the supplied data to defend their explanation once they had chosen it while the individuals in the higher performing groups used the data to distinguish between explanations or to confirm the accuracy of an explanation *before* making a final decision about which explanation best accounted for the discrepant event. This strategy appears to have helped the higher performing groups produce a more accurate and convincing argument. Although other researchers have found that many students do not use data in an appropriate manner to generate explanations or to support their claims (Kelly et al. 1998; McNeill et al. 2006; Sandoval and Millwood 2005), this finding suggests that some students privilege the use of data as a way to evaluate alternative explanations and others do not. This is a relatively new finding in this context and another contribution of this study.

While we have presented these differences sequentially as distinct characteristics, these differences appear interrelated. For example, individuals in the two lower performing groups introduced fewer ideas to the discussion than the individuals in the two higher performing groups. These groups also accepted ideas outright more frequently and challenged ideas less frequently through discussion. This suggests that the individuals in lower performing groups agreed with the ideas that were proposed by other group members and, as a result, these individuals did not critically evaluate these ideas or discuss potential alternatives as a group. Similarly, the individuals in the lower performing groups apparently did not understand the importance of using the supplied data to assess the merits of an idea. Consequently, these individuals did not use the available data to evaluate the alternative explanations or critique the ideas of other group members. The lower performing groups, apparently as a result, accepted ideas outright (e.g., “that makes sense” or “that sounds good to me”) during their conversations more frequently and only used the supplied data only to defend their final explanations. Thus, interrelationships between these differences seem likely.

Limitations

It is important to note that the need for experimental control over variables such as the types of explanations students evaluated and the data they had available in the original study resulted in a very specific task for the student to complete. As discussed earlier, there are a number of different ways to promote scientific argumentation inside the classroom and the task we used in this study represents but one of many. It is therefore important to keep in mind that the results of this study may have been different if the nature of the task were changed. For example, more comments that focused on the co-construction of knowledge (e.g., summarizing, revising, supporting, or adding to the ideas of others.) and fewer oppositional comments may have led to a better written argument during a task that requires 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 study might have also been different if we had chose to explore a different group outcome (e.g., a more sophisticated individual understanding of a topic or the processes involved in scientific inquiry). The results of the current study therefore call for further replication with different types of tasks and with different types of desired outcomes to support broader generalization and specification.

It is also important to underscore that we used an emergent and iterative analytical approach in this study to facilitate the identification and documentation of group interactions and processes that seem to promote or constrain the development of a higher quality written scientific argument. Although the nature of this type of analysis, the small sample size, and the specifics of the task limits broad claims of generalizability, the findings reported here can lay the groundwork for future investigations that incorporate larger samples or experimental designs. These findings can also be used to inform the development of new analytical frameworks or assessment instruments that can be used to examine the nature of collaborative scientific argumentation that takes place within the classroom. With these issues in mind, we will first highlight several factors that may have contributed to the observed differences in ways the high and low performing groups engaged in collaborative scientific argumentation. We will then discuss the implications of this research for the teaching and learning of science and provide several recommendations for the development of new curricula, instructional models, and analytical frameworks.

Potential Factors that May Have Contributed to Observed Differences in Group Interactions

One potential factor that might explain the observed differences in the ways higher and lower performing groups engaged in collaborative scientific argumentation is the students' lack of familiarity with the culture of western science. As noted at the beginning of this article, the work of scientists is shaped by the shared goals, norms, and epistemological commitments of the scientific community. These aspects of the culture of science influence the actions of scientists and how scientists think, interact, and communicate with each other. Students who are less familiar with the culture of western science are therefore less likely to understand how the nature of scientific argumentation is different from the nature of the argumentation in other contexts or how scientists rely on different habits of mind (e.g., skepticism, openness to new ideas, appreciation of empirical evidence, etc.) or norms to govern how they do their work. This seemed to be the case for the individuals in the lower performing groups. These individuals seemed to interact with each other, new ideas, and the available materials in a manner that was not well aligned with the goals, norms, and commitments of the culture of science (or at least the idealized western conception of it) whereas the individuals in the higher performing groups seemed to use modes of interaction and thinking that seemed to better reflect the shared and accepted traditions of this culture.

A second potential factor that may explain the observed differences in the interaction patterns of the higher and lower performing groups in this context is the students' lack of familiarity with genuine collaboration in the classroom. Thus, the "unwritten rules" (Lemke 1990) of schools and teachers may have had a negative impact on the ways some of the students 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 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 and Inagaki 1991) or break up tasks into smaller pieces which individuals can accomplish on their own (Cohen 1994; Eichinger et al. 1991) because these strategies are often more efficient or useful in the classroom. Although these strategies can be valuable during some types of group work, using these types of strategies during a task designed to engage students in collaborative scientific argumentation may have prevented the members of the lower performing groups from generating a high quality written argument. This means that the culture of the classroom and the modes of interaction or thinking that are valued and encouraged inside a classroom by the teacher might also hinder some students from engaging in collaborative scientific argumentation in a manner that reflects our theoretical perspective.

Finally, a number of studies (see Carlsen 2007 for a review) have shown that cultures and social groups differ in the degree to which they privilege or encourage specific modes of communication and types of interaction. As Carlsen explains, productive communication requires students to know how "to take turns without the teacher's directions, how to hold (and yield) the floor, and how to make sense to (and of) others" (Carlsen 2007, p. 64), which can be "complicated by speaker differences of gender, culture, ethnicity, and so on" (p. 64). Some of these modes of communication and interaction are well aligned with those valued and encouraged in a western conception of scientific argumentation and some are not. The work of Brown and Ryoo (2008) also supports this notion. Their research indicates that discourse (i.e., language in all of its semantic and representational forms) and language use (i.e., a collection of words that are used to represent ideas) varies by community, context, and is often influenced by the identities (both perceived and projected) of the participants. Their research also suggests that students need to have a shared cultural understanding of science in order to communicate with the specialized 'D' discourse (Gee 1999) or 'non-vernacular' language of science. Thus, the observed differences in the nature of the discussion may have reflected existing interpersonal dynamics, cultural norms (which include, but are not limited to, the culture of students' homes, communities, peer groups), individual identities, or individual personalities in terms of the students' familiarity and comfort with the version of scientific argumentation introduced to them through this curriculum.

These factors, either acting in isolation or in unison, might have had a substantial influence on how students engage in collaborative scientific argumentation. These factors, for example, would explain why the individuals in the higher performing groups questioned their peers' ideas or positions more often and were more likely to voice disagreement, ask for a justification or to offer a critique than the individuals in the lower performing groups. Students in the higher performing groups might have been simply more comfortable with this type of discourse than the students in the lower performing groups. These factors might also help explain why the students in the higher performing groups used different criteria to evaluate ideas than the students in the lower performing groups. These criteria, which were aligned more closely with the introduced framework and the culture of science, might have seemed more familiar or made more sense to the students in the higher performing groups because there was less disconnect between these criteria and the ones these students use in other contexts. Although these factors are only conjectures, we believe that targeted research that examines these disconnects is needed and might help us better understand the underlying reasons for inter-group variation in performance.

Implications and Recommendations

The findings reported here, given the context and the nature of intervention, suggest that science educators will need to mindful of the challenges that occur at the intersection of collaborative work and scientific argumentation. For example, in order to collaborate with each other, students need to be willing to voice ideas, discuss ideas when they are mentioned, and hold each other accountable. Some students might find these types of interactions unfamiliar or uncomfortable and therefore struggle when asked to collaborate with others. Similarly, even when a group of students is able to engage in genuine collaboration, a lack of familiarity with the goals and norms of scientific argumentation might also hinder their ability to reach a desired group outcome. For example, students may not use criteria valued in science or use data only to support their ideas rather than to evaluate alternatives. Both of these issues seemed to hinder the performance of the two lower performing groups in this study. Overall, these findings indicate that efforts to help students learn how to engage in collaborative scientific argumentation will require a more explicit focus on the ways students interact with each other, ideas, and available materials and not just on the ways students propose, support, and challenge claims.

In order to help students produce better group outcomes though collaborative scientific argumentation, we suggest that science educators focus their attention on helping students learn to view ideas as objects of cognition (Kuhn 1993; Mason 2001) and encourage students to adopt and use different criteria and strategies for evaluating or supporting ideas than they apply in everyday contexts. The argument framework that we used in this study, which not only defines what counts as a good argument in science but also highlights criteria that students can use to evaluate the quality of an explanation or an argument, may be a fruitful option for this type of endeavor. This framework not only provides guidance for students as they work but also makes several indicators of quality explicit. This framework may also help to alleviate some of the guesswork and uncertainty that is often associated with learning a new practice and help align classroom practice with the culture of western science for both students and teachers.

We conjecture that this type of focus will not only improve student and teacher understanding of the goals and processes of scientific argumentation but will also facilitate the students' ability to cross the borders (Aikenhead 2001) between the cultures of western science and those typical of school and their everyday lives. Although this is only speculative, this type of focus should help students learn more from (i.e., content) and about (i.e., practices) scientific argumentation. This focus, however, will require scientific argumentation to take a more central role in the teaching and learning of science and it will require skilled and knowledgeable science teachers who understand the goals and processes of scientific argumentation and how these goals and processes differ from the nature of argumentation that takes place in other contexts. It will also require science teachers to adopt and use instructional approaches that are more knowledge-, learner-, and community-centered (Donovan and Bransford 2005).

This study also provides a potential foundation for researchers interested in developing new analytical frameworks or assessment instruments that can be used to examine the nature or quality of the collaborative scientific argumentation that take place inside the classroom. The results of this study, for example, indicate that the higher performing groups voiced more ideas, were more likely to discuss ideas when they are proposed, and challenged or critiqued ideas more often. Science educators can look for these types of interactions and then use this information to evaluate overall group performance. Similarly, the results of this study indicate that science educators will need to be more mindful of the types of criteria that students are using to support or evaluate the merits of an idea and how a group is choosing to use available data as they work. All five of these aspects of collaborative scientific argumentation seem to influence group performance when students are required to evaluate several alternative explanations for a discrepant event and then craft a written argument. In order for teachers and researchers to support students, therefore, we recommend the development of new analytical and assessment instruments that focus on all five of these important aspects of collaborative scientific argumentation in addition to the more traditional foci (e.g., use of data or rebuttals) so that researchers can better explore how students evolve in their understanding of these complex practices and teachers can focus their thinking, teaching, and evaluation to better support students learning to engage in these complex practices.

Final Thoughts

The current study demonstrates that the development of a high quality written argument through collaborative scientific argumentation requires more than simply pooling knowledge or combining different cognitive and monitoring resources. In other words, a group outcome cannot be viewed simply as the sum of individual abilities. How individuals interact and the decisions they make while engaged in this process have profound influences on group meaning making and the nature of the arguments groups generate. It is our hope that the findings of this study can inform the design of future studies and the development of new analytical frameworks that can be used to examine the nature and causes of inter-group variation in performance during collaborative scientific argumentation. Once we understand more about the underlying causes for the differences we see in the performance of groups, we should be able to more effectively structure activities in the classroom to help student learn more from and about collaborative scientific argumentation.

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: American Association for the Advancement of Science.

Aikenhead, G. (2001). Student's ease in crossing cultural borders into school. *Science Education, 85*, 180–188.

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

Andriessen, J., Baker, M., & Suthers, D. (2003a). 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: Kluwer.

Andriessen, J., Erkens, G., Van de Laak, C., Peters, N., & Coirier, P. (2003b). 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). The Netherlands: Kluwer.

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

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

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

Boulter, C. J., & Gilbert, J. K. (1995). Argument and science education. In P. J. M. Costello & S. Mitchell (Eds.), *Competing and consensual voices: The theory and practices of argument*. Clevedon: Multilingual Matters Ltd.
Brown, B. A., & Ryoo, K. (2008). Teaching science as a language: a "content-first" approach to science teaching. *Journal of Research in Science Teaching, 45*(5), 529–523.
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.
Carlsen, W. S. (2007). Language and science learning. In S. K. Abell & N. Lederman (Eds.), *Handbook of research on science education*. Mahwah: Erlbaum.
Cartier, J. L., & Stewart, J. (2000). Teaching the nature of inquiry: further developments in a high school genetics curriculum. *Science and Education, 9*(3), 247–267.
Chi, M. T. (1997). Quantifying qualitative analyses of verbal data: a practical guide. *The Journal of the Learning Sciences, 60*(3), 271–315.
Clark, D. (2006). Longitudinal conceptual change in students' understanding of thermal equilibrium: an examination of the process of conceptual restructuring. *Cognition & Instruction, 24*(4), 467–563.
Clark, D., & Sampson, V. (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.
Cohen, E. G. (1994). Restructuring the classroom: conditions for productive small groups. *Review of Educational Research, 64*, 1–35.
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.
Donovan, M. S., & Bransford, J. (Eds). (2005). *How students learn: Science in the classroom*. Washington: National Academy Press.
Driver, R., Asoko, H., Leach, J., Mortimer, E., & Scott, P. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher, 23*, 5–12.
Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84*(3), 287–313.
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: Springer.
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.
Erduran, S., Osborne, J., & Simon, S. (2004a). The role of argument in developing scientific literacy. In K. Boersma, O. deJong, H. Eijkelhof, & M. Goedhart (Eds.), *Research and the quality of science education*. Dordrecht: Kluwer.
Erduran, S., Simon, S., & Osborne, J. (2004b). TAPping into argumentation: developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education, 88*, 915–933.
Gee, J. (1999). *An introduction to language analysis: Theory and method*. New York: Routledge.
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: American Psychological Association.
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.
Hogan, K., Nastasi, B., & Pressley, M. (2000). Discourse patterns and collaborative scientific reasoning in peer and teacher-guided discussions. *Cognition and Instruction, 17*(4), 379–432.
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: 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.

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

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

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.

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

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: Ablex.

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: Erlbaum.

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: Erlbaum.

Mason, L. (2001). Introducing talk and writing for conceptual change: A classroom study. In L. Mason (Ed.), Instructional practices for conceptual change in science domains. *Learning and Instruction, 11*, 305–329.

McNeill, K., & Krajcik, J. (2008). Scientific explanations: characterizing and evaluating the effects of teachers' instructional practices on student learning. *Journal of Research in Science Teaching*, *45*(1), 53–78.

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *The 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: 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.

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: Erlbaum.

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. *The 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: National Center for Education Statistics.

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

Sadler, T., Barab, S., & Scott, B. (2007). What do students gain by engaging in socioscientific inquiry. *Research in Science Education, 37*, 371–391.

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

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

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.

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). The Netherlands: Kluwer.

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

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

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: 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: The National Academies Press.

Suthers, D., & Hundhausen, C. (2001). Learning by constructing collaborative representations: An empirical comparison of three alternatives. In P. Dillenbourg, A. Eurelings, & K. Hakkarainen (Eds.), *European perspectives on computer-supported collaborative learning* (pp. 577–592). Maastricht: University of Maastricht.

Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.

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

Veerman, A. (2003). Constructive discussions through electronic dialogue. In J. Andriessen, M. Baker, & D. Suthers (Eds.), *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments* (pp. 117–143). The Netherlands: Kluwer.

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

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.