JOURNAL OF COGNITION AND DEVELOPMENT, 11(3):293–327

ISSN: 1524-8372 print/1532-7647 online
DOI: 10.1080/15248372.2010.485335

Psychology Press
Taylor & Francis Group

Developing Argument Skills Across Scientific and Social Domains

Kalypso Iordanou
Columbia University

Instant-messaging software was used as a method to promote development of argumentation skills in middle schoolers. Transfer of skills across content domains was the major question investigated. Forty sixth graders engaged in electronic dialogues with peers on a controversial topic—for half a science topic (dinosaur extinction) and for half a social topic (home schooling). During 13 sessions, participants worked with a partner in arguing with a succession of pairs of classmates who held an opposing view on the topic; in addition, they engaged in some reflective activities based on transcriptions of the dialogues. Another 18 sixth graders served in a control (nonintervention) condition. Although transfer occurred in both directions, science condition participants exhibited transfer of skills to the social topic to a greater extent than did social condition participants to the science topic. Results show the transfer, and hence generality, of developing argument skills but also suggest the importance as well as feasibility of fostering argument skills within science and social domains.

Contemporary authors on education have repeatedly expressed the view that students must acquire strong, flexible, higher-order thinking skills if they are to be equipped for life in the 21st century (Bereiter, 2002; Kuhn, 2005; Wagner, 2008). Yet, the extent of empirical research in the field of cognitive development needed to support these arguments has been limited, especially relative to the amount of research devoted to cognitive development in the early years of life. There has been a fair amount of research on the development of scientific reasoning (see Lehrer & Schauble, 2006; Zimmerman, 2000, 2007, for reviews), but key questions such as the degree

Correspondence should be sent to Kalypso Iordanou, Teachers College, Columbia University, 525 West 120th Street, Box 118, New York, NY 10027, USA. E-mail: kai2101@columbia.edu

to which educational interventions can promote this development and the most effective methods for doing so remain unresolved (Dean & Kuhn, 2007; Klahr, 2000; Kuhn, 2009; Kuhn & Dean, 2005; Kuhn, Iordanou, Pease, & Wirkala, 2008; Strand-Cary & Klahr, 2008). Until recently, most research on scientific reasoning has been devoted to control of variables, and much less attention has been paid to skills of argumentation, despite their centrality in both science and social, and academic and nonacademic, contexts (Kuhn, in press). Indeed, the case has been made by Oaksford, Chater, and Hahn (2008) that argumentation is the umbrella under which all reasoning lies; in their words, it is "the more general human process of which more specific forms of reasoning are a part" (p. 383).

The available developmental research indicates that children and adolescents do not exhibit strong argument skills (Klaczynski, 2000; Kuhn, 1991; Means & Voss, 1996; Perkins, 1985; Stanovich, Toplak, & West, 2008). Moreover, little improvement is observed with age, and skill in informal reasoning is not easy to induce. In his studies of belief bias in reasoning, Klaczynski and colleagues (Klaczynski; Klaczynski & Gordon, 1996) have shown that extrinsic motivation has some but only a limited effect; as is the case with other cognitive skills, development of meta-level understanding—in particular, enhanced understanding of discourse goals in the case of argumentation (Kuhn, Goh, Iordanou, & Shaenfield, 2008)—and development of effective strategies to meet these goals are essential. Consistent with this claim, strong argument skills have been associated with more mature epistemological understanding (Kuhn, 1991). In particular, Klaczynski and Lavallee (2005) found epistemological beliefs—such as open-mindedness and need for cognition—to be negatively correlated with belief-based reasoning.

The study presented here follows a line of work devoted to fostering the development of argument skills in early adolescents based on engagement and practice in argumentation (Felton, 2004; Kuhn, Shaw, & Felton, 1997; Kuhn, Goh, et al., 2008; Kuhn & Udell, 2003; Udell, 2007). Dialogic argument that takes place between people, in addition to its importance in its own right, is a promising pathway for the development of the nondialogic, individual expository argument skills that figure prominently in academic contexts. The two are intricately connected (Billig, 1987; Graff, 2003; Kuhn, 1991), with dialogic argument providing the "missing interlocutor" (Graff) that is lacking in the individual expository argument of a single individual. Moreover, dialogic argument has the advantage of building on the familiar form of everyday conversational exchange.

The intervention method used in the work presented here is modeled on the engagement and practice method first implemented by Kuhn et al. (1997) and extended by De Fuccio, Kuhn, Udell, and Callender (2009), Felton (2004), Kuhn and Udell (2003), Kuhn, Goh, et al. (2008), and Udell

(2007). In these studies, young adolescents engage in dialogic argument with a succession of peers who hold an opposing view on a social issue. Initially, this work shows, adolescents focus their efforts on exposition of their own position, consistent with the "myside bias" reported by Perkins (1985) and others (Baron, 1995; Stanovich & West, 2007). They neglect attending to the opponents' claims and attempting to weaken their force. Walton (1989) identifies this objective, along with securing commitments to support one's own claims, as the two basic goals of skilled argumentation. With sustained practice, participants in these interventions gradually attend more to the other's contributions to the dialogue and develop skill in addressing them directly in the form of counterarguments that attempt to weaken their force. In addition, they show increased metacognitive control over the dialogic process (Kuhn, Goh, et al., 2008). The key question the present study addresses is that of transfer of argumentation skills across domains. Do skills developed in one domain transfer to others? The answer to this key question has practical implications, as well as theoretical ones with regard to the nature of the skill itself. Can developing skill in argumentation itself be identified, or is such skill specific to the content and/or context in which it is acquired? Although there is evidence for such transfer between social and scientific domains in scientific reasoning (Kuhn, Garcia-Mila, Zohar, & Andersen, 1995; Zimmerman, 2000, 2007), the question has not been specifically addressed with respect to argumentation skills.

A particular feature of the present study is its use of instant-messaging (IM) computer software as the medium of discourse, following the successful use of this method by Kuhn, Goh, et al. (2008). Several studies by other researchers (Andriessen, 2006; Bell & Linn, 2000; Clark & Sampson, 2007; see Clark, Stegmann, Weinberger, Menekse, & Erkens, 2008, for a review of studies using technology-enhanced environments to support argumentation) suggest that this medium is a fruitful one for scaffolding argumentation in science domains. In contrast to these studies, however, the present method involves no software-based scaffolding of argument construction and evaluation, beyond the IM software itself. As a rationale for use of the IM methodology, Kuhn, Goh, et al. (2008) suggest that it offers support for the development of meta-level awareness regarding the discourse. It has the benefit of providing an immediately available, permanent record of the discourse for participants to reflect on, in contrast to the conditions of real-time verbal discourse, where the contents of each contribution to the dialogue immediately disappear as soon as they are spoken.

Research on adolescents' argumentation skills has until now focused on their reasoning about social issues. The present study includes a topic in the physical science domain. The choice of a physical science domain as the additional domain that allows investigation of the question of transfer was prompted by the growing centrality of argumentation in the field of science education. Science educators increasingly have come to see argumentation as central to science learning (Jiménez-Aleixandre & Erduran, 2008). Yet, studies that have undertaken explicit teaching of argument skills in a scientific context have shown mixed results (Mercer & Littleton, 2007; Osborne, Erduran, & Simon, 2004; Zohar & Nemet, 2002), specifically with respect to key skills of considering alternative positions and integrating evidence with claims. Zohar and Nemet's findings in particular suggest that engagement and practice in discourse itself is essential to developing such skills, as do related studies by Mason (1998) and Naylon, Keogh, and Downing (2007).

A potential problem with applying in a science domain the dialogic methods used in work to foster development of argument skills is that young students are widely regarded as lacking sufficient knowledge about science topics to engage in productive debate. In the present work, this challenge is addressed by providing participants with a constrained knowledge base (a set of "possibly relevant facts") that is equated across participants (and topics) and can serve as a resource for their argumentation.

The research design is a straightforward one in which participants are randomly assigned to one of two intervention conditions—social content or science content—and their argument skill level is assessed before and after the intervention in both the social and science domains. A third non-intervention (control) group is included for comparison. The key question asked here is the degree to which advances in argument skill within the intervention experience situated in a social-domain topic will transfer to skill with respect to the science-domain topic, and vice versa. Does such transfer occur, and if so does it occur in a symmetrical manner (i.e., equally in either direction)? Or is the pattern asymmetrical, with transfer in one direction (science to social) easier or more difficult to achieve than transfer in the other direction (social to science)?

In addition, a measure of epistemological understanding (Kuhn, Cheney, & Weinstock, 2000), described more fully in the "Initial and Final Assessments" section, was administered at the beginning and end of the intervention, to probe how the intervention may have affected epistemological understanding in scientific and social domains. A number of researchers concerned with argumentation (Hofer & Pintrich, 2002; Kuhn, in press; Mason, 2003) have proposed that as important as advances in argument skill are advances in epistemological understanding of the purpose and goals of argument in the construction of knowledge. In particular, previous work (see Greene, Azevedo, & Torney-Purta, 2008; Kuhn, in press; Kuhn, Iordanou, et al., 2008; Mason & Scirica, 2006; Muis, Bendixen, & Haerle, 2006) identifies two major transitions that have the potential to be affected by an argumentation intervention. One is the transition from an absolutist level of belief in certain, accumulating knowledge to a multiplist level in which the human, subjective, and constructive aspect of knowing dominates. A second transition is from a multiplist to an evaluativist level of epistemological understanding, in which the subjective and objective components of knowing are integrated and it is recognized that claims can be compared and evaluated in a context of evidence and argument (Kuhn et al., 2000).

METHOD

Participants

Participants were 62 sixth graders from a public elementary school in a middle-class suburban area in the country of Cyprus. The 44 participants in the experimental conditions consisted of the entire sixth grade, and the 18 participants in the control condition were randomly chosen from the following year's sixth graders at the same school. All were 11 or 12 years of age; 38 were boys and 24 were girls (27 boys and 17 girls in the experimental condition and 11 boys and 7 girls in the control condition). Participants were primarily from a middle-class population. Roughly 30% were from minority ethnic groups. Four participants in the experimental condition (3 boys and 1 girl), whose language abilities were judged by the school system as needing remediation, participated but were not included in the analysis.

Initial and Final Assessments

Participants' argument skills were assessed on the home-school (social) topic and the dinosaur extinction (science) topic, described in the "Individual Argument" section, at both initial and final assessments. Control group participants underwent assessments on both topics at the same times of year—with the same time interval between initial and final assessment—as the experimental group, but a year later. At both assessments, participants were also administered two items—one in a social domain and one in a science domain—from Kuhn et al.'s (2000) instrument assessing levels of epistemological understanding. Each item involved two conflicting positions on an issue. Participants were asked whether only one position could be right (identified by Kuhn et al. (2000) as an absolutist epistemological level) or both could have some rightness. If participants chose the latter option, they were then asked whether the positions were necessarily equally valid (a multiplist or relativist level) or whether one position might be better or more right than the other (an evaluativist level).

Participants were told they were going to work on a project regarding some issues that have come up and must be resolved in forming a new town in an unspecified location. The issues they would debate, they were told, involved the school that would be established in the town. One issue was whether children in the town must attend the town school or whether parents can be allowed to home school their children if they wish to (social topic). The other issue concerned which of two competing explanations should be presented in science classes regarding dinosaur extinction (science topic).

1. Individual argument. Initial positions and supporting arguments regarding the social topic—home-school (HS)—and the science topic—dinosaur extinction (DE)—were assessed individually in writing. A short passage introduced the scenario. For the HS topic, the scenario was the following:

> A problem has come up. The Ito family has moved to the edge of town from far-away Japan with their 11-year-old son Aki. Aki was a good student and baseball player back home in Japan. The family will stay in the new town for 1 year, and Aki's parents have decided that in this new place, they want to keep Aki at home with them and teach him in Japanese, instead of having him be at the school. The family speaks only Japanese, and they think Aki will do better if he sticks to his family's language and doesn't try to learn Greek. They say they can teach him everything he needs at home. What should happen? Is it okay for the Ito family to live in the town but keep Aki at home, or should they send their son to the town school like all the other families do?

Participants indicated their position by choosing among the following options: "Aki should go to the town school," "Aki can be taught at home," and "Undecided."

For the DE topic, the scenario was the following:

> A chapter is going to be included in the sixth graders' science book to explain dinosaurs' disappearance. A problem has come up. A scientist, who was called to give his view of the issue, claims that dinosaurs were quickly exterminated due to the difficult climate conditions caused by the collision of an asteroid with the Earth. However, another scientist who was also called to give his view about the issue claims that dinosaurs gradually disappeared due to the difficult climate conditions caused by a series of giant volcanic eruptions that lasted for several million years. What should happen? Should students be taught that dinosaurs were quickly exterminated by the collision of an asteroid with Earth or that dinosaurs gradually disappeared due to giant volcanic eruptions?

Response options for this topic were: "Dinosaurs were quickly exterminated by the collision of an asteroid with the Earth," "Dinosaurs gradually disappeared due to giant volcanic eruptions,'' and ''Undecided.'' For each scenario, participants were also asked to indicate the certainty of their position on a 6-point Likert scale, with endpoints labeled ''totally certain'' and ''totally uncertain.'' They were then asked for reasons supporting their position and finally for reasons that would support the opposing view.

2. Dialogic electronic argument with opposing-view partner. For each topic, two groups of 20 participants each were formed. The assignment was based on the position statement expressed on the 6-point opinion scale in the initial individual assessment, except for a few cases where it was not clear and it was necessary to consider the reasons participants offered in order to make the assignment. The participants who indicated they were undecided gave reasons on both sides of the issue. They were assigned to one or the other position in a way that served to equate the number of participants on each side. For the DE topic, one group consisted of 19 participants who chose the volcano position and 1 undecided participant; another group consisted of 12 participants who were in favor of the asteroid position and 8 who were undecided. Similarly, for the HS topic, two groups of 20 participants each were formed. One group consisted of 20 participants who chose the home-school option, and another group consisted of 15 participants who chose mandatory town-school and 5 who were undecided. In the control group, no participants were undecided, and by chance, there was an equal number of participants supporting each side for both topics.

By drawing one from each of the contrasting groups for the topic, pairs of opposing-side participants were formed for each topic. Pairs were different for the two topics. This pair engaged in an initial and final dialogic interchange on the topic, prior to and following the intervention. To conduct these dialogues, the two members of the pair were situated on different sides of the room in a computer lab, facing away from one another and thus restricting verbal exchange or eye contact. Dialogues were implemented by IM chat software (MSN). Before pairs discussed each topic, they were reminded of the scenario and given the following instruction:

> You are here to have a serious discussion about the topic of what caused dinosaurs' extinction (or home-school). You have different views on the topic. State your views and your reasons for them. Find out where you agree and disagree. If you disagree, figure out why and try to reach an agreement if you can.

The dialogues lasted up to 20 minutes; participants completed the dialogue when the time elapsed or earlier if they said they had finished. Each participant engaged in a dialogue on the social topic with an opposing-side partner and in a dialogue on the science topic with a different opposing-side partner.

The order of dialogues on the science and social topics was counterbalanced across participants. Transcripts of the dialogues were saved for analysis.

Intervention

Each participant in the experimental condition was randomly assigned to one of two conditions: a) the social condition (SOC) or b) the science condition (SCI). Those in the SOC participated in an intervention involving the HS topic. Those in the SCI participated in an intervention involving the DE topic. The two genders were equally represented in the two conditions. The two intervention conditions were identical except for the topic (HS or DE). Participants in the control condition did not participate in an intervention.

The intervention took place during thirteen 40-minute sessions occurring twice per week in the participants' classroom. Because of school holidays, it took approximately 2.5 months to be completed. The two experimental interventions (SOC and SCI) took place simultaneously, with participants taking part in only one of the two. Participants were told that they were preparing for a final "showdown" in which they would debate their topic, either HS or DE, with the group of their classmates who held the opposing view.

1. Preparation for supporting reasons with evidence. Following the pedagogical objectives identified by Kuhn and Udell (2003; Appendix A), an initial goal was to make explicit the concept of evidence as strengthening a claim. Participants in each condition were divided into the two teams based on their position and used in formation of pairs for initial and final assessments, as described earlier. (Participants in the SOC were divided into the home-school and the town-school teams. Participants in the SCI were divided into the volcanoes and the asteroid teams). Participants were given a list of "some possibly relevant facts." This list contained 16 facts, 8 supporting each position, presented in a random order. An illustration of 1 from the DE fact sheet is, "Large quantities of iron and other metals that include iridium have been found at the Earth's core." An illustration of 1 from the HS fact sheet is, "There are published curriculum books available in bookstores that guide the teaching of subjects like math and history. They suggest what to teach the child at each point." Participants were asked to review this information individually and then to decide as a team if there were any facts they wished to make use of. They discussed as a group what the implications of each of these facts were. An adult coach, who was a researcher, facilitated each group's work, helping to keep them focused on their task when necessary and answering questions but offering no direct instruction or feedback.

2. Paired dialogic electronic argument with opposing-view pair. Same-gender pairs (who shared the same view on the topic) were formed within each team. The same-side pairs remained together until the showdown preparation (see point 4 below). The pair conducted an electronic dialogue with another pair on the opposite side of the room who held the opposing position on the topic. As was the case for the dialogic electronic argument at initial and final assessments, participants did not know the identity of their IM opposing partners. Oral instructions provided to each pair were to collaborate with their partner to determine what they wished to say and, when they reached agreement, to enter their response and send it to the opposing pair. Two adult coaches circulated, and when asked, they helped with technology issues and especially the less skillful typing pairs to type their responses. Dialogues lasted an average of 25 minutes. At the next session, each pair debated with a different opposing pair until each pair had debated every opposing pair—a total of five paired dialogue sessions.

3. Reflective analysis of transcripts from previous argument sessions. After three dialogue sessions had been completed, reflective analysis was introduced. In this activity, a pair analyzed the printed transcript of their immediately preceding session's dialogue. Two reflection sheets were provided: the "Other Argument" and the "Own Argument" reflections. With the help of the "'Other Argument' Reflection Sheet" (Appendix B), the pair's task was to analyze the opposing side's argument and reflect on the effectiveness of the counterargument they made and consider possible improvements to this counterargument.

With the help of the "'Own Argument' Reflection Sheet" (Appendix C), the pair's task was to review and evaluate the counterarguments made by the opposing side to their own arguments and their rebuttals to these counterarguments, and consider possible improvements to their rebuttals. When some pairs finished the reflective analysis of their own dialogue's transcript, they exchanged transcripts and reflection sheets with other pairs to give and receive feedback. The coaches facilitated the process by prompting participants to think of whether their responses to the opposition were as convincing and effective as they could be.

The first reflective session, using the "Other Argument" reflection sheet, occurred after the third dialogic session; the second reflective session, using the "Own Argument" reflection sheet, occurred after the fourth dialogic session; the third reflective session, using both reflective sheets, occurred after the fifth dialogic session (see Appendix D for design summary).

4. "Showdown" preparation session. The participants who had been working together as a pair for dialogue and reflection sessions separated and were assigned to two different preparation teams. One team was assigned to be "own argument" specialists and the other was assigned to be "other argument" specialists. Each preparation team had an adult coach to facilitate the group process. Both groups were told that the purpose of this session was to prepare for the impending "showdown."

The "own argument" specialists were told that their task was to become familiar with the possible counterarguments the opposition might assert and to prepare rebuttals to use in the showdown. The team created a set of "own argument–counter–rebuttal" sequences that were recorded onto color-coded cards, distinguishing each part of the argument sequence. The reflection sheets completed in previous sessions were made available for this activity and further possible improvements were considered.

Members of the other team were the "other argument" specialists. Their task was to review effective counterarguments to use when faced with opponents' arguments. The cards produced by this team reflected the argument sequence of "other argument–counter." Again, the reflection sheets were made available for this activity and further possible improvements were considered.

5. "Showdown". Participants on each side of the issue were divided into two teams of five members—Team A and Team B. The previous "specialists" (own argument and other argument) were represented equally on Team A and Team B.

Team A and Team B participants on each side were seated in different rooms, and the two sides communicated through IM software. The dialogue was projected onto a wall screen in each room. All members collaborated to come to an agreement on the text to be sent to the opposing side. One member of each team was designated as typist. During the first half of the showdown, the A teams debated. At half-time, a team change took place and the B teams continued the debate. The showdown thus consisted of a single electronic dialogue between the two sides of approximately 40 minutes duration.

6. Judging and feedback. The electronic dialogue produced in the showdown was represented in an argument map prepared by the researchers. Different columns appeared for each team, with their contributions arranged in order of occurrence from top to bottom. All statements were represented and connected by lines to show their interrelation. Different colors were used to label statements as effective, ineffective, or neutral argumentative moves. A point system was also applied, making it possible to declare a winning team. The argument map and associated point scoring were presented to participants in a session following the showdown.

Post-Intervention Assessment

The final assessment was identical to the initial assessment. Participants engaged in a single computer-mediated dialogue with the same partners as in the initial assessment on both the HS and DE topics, as described under "Initial Assessment." Because two participants were absent during the final assessment, their partners at initial assessment (four participants, two for each topic) were matched with another participant who already had participated in a dialogue.

RESULTS

To examine the effectiveness of the intervention in developing argumentation skills in topics that spanned two domains—physical science or social science—and to investigate whether argumentation skills developed in one domain transfer to another, electronic dialogues at initial and final assessments were examined.

Coding Electronic Discourse Strategies

The analysis is based on the 116 electronic dialogues produced at initial and final assessments by participants in the two experimental and one control conditions on the social and the science topics (56 dialogues per topic). Two participants, one from the social experimental condition and one from the control condition, were absent during the final assessment and were excluded from the analysis. The dialogues were analyzed based on the argumentive discourse scheme used in previous research (Kuhn, Goh, et al., 2008).

The coding scheme is a functional one, designed to assess the functional relation between an utterance and the opponent's immediately preceding utterance. An utterance is defined as the minimum idea unit that serves a specific function in the conversational exchange. Each utterance in the discourse was segmented and categorized as reflecting one of the single-utterance argumentative operations in the coding scheme (Appendix E).

For coding meta-level statements—that is, statements about the dialogue rather than contributions to it—the revised coding scheme developed by Kuhn, Goh, et al. (2008) was used. Each meta-level utterance in the discourse was segmented and categorized as reflecting one of the single-utterance meta-level operations in the coding scheme (Appendix E).

Thirty percent of the dialogues were randomly selected and used to calculate inter-rater reliability. Two trained coders blind to the treatment, time, and identity of the participants, participated in segmenting and coding. Coders' percentage of agreement on coding was 89% (Cohen's Kappa = .87). After establishing inter-rater reliability, the remaining electronic dialogues were segmented and coded by one of the raters, again blind to treatment, time, and identity of participants.

Length and Number of Utterances

Before turning to the main analysis, based on type of argumentation strategies used, two quantitative indicators were examined that, while hardly conclusive on their own, provide an initial, albeit superficial, indication of change in argument skill. These are changes in the number of utterances in a dialogue and in the length of those utterances. These analyses showed that participants' contributions to the dialogues became both longer and more numerous from initial to final assessment. A 3 (Condition) $\times$ 2 (Time) $\times$ 2 (Topic) repeated-measures analysis of variance (ANOVA) comparing the three conditions was used to assess whether the conditions had differential effects on the number of words contained in each utterance. A three-way analysis revealed a three-way Topic $\times$ Time $\times$ Condition interaction, $F(2, 53) = 8.50$, $p = .001$, partial $\eta^2 = .24$. On the social topic, SOC participants doubled the length of their utterances—from 6.71 ($SD = 2.15$) to 12.66 (5.06)—by the end of the intervention, whereas SCI participants showed no change—from 8.39 (2.65) to 8.93 (3.18). On the science topic, both experimental conditions showed increases in the length of their utterances, from 5.77 (1.94) to 9.56 (4.22) in the case of the SOC and from 6.81 (2.01) to 10.09 (4.22) in the case of the SCI. Participants in the control condition showed a decline in length of utterances on both topics, from 9.05 (4.15) to 6.57 (1.79) on the social topic and from 7.99 (2.32) to 5.62 (1.51) on the science topic.

A three-way analysis of mean number of coded utterances revealed a significant three-way interaction, $F(2, 53) = 4.66$, $p = .014$, partial $\eta^2 = .15$. SOC participants increased from 9.70 ($SD = 3.07$) to 11.90 ($SD = 3.43$) on the social topic and from 8.40 (3.39) to 12.20 (4.31) on the science topic. SCI participants increased from 8.45 (4.42) to 14.75 (5.84) on the social topic and from 7.95 (2.89) to 12.10 (3.98) on the science topic. Control condition participants, in contrast, decreased from 9.80 (4.70) to 4.90 (1.87) on the social topic and from 8.80 (3.68) to 6.50 (4.54) on the science topic. Between-subjects analysis, using a Helmert contrast, showed that the two experimental conditions differ from the control condition, $F(2, 53) = 9.67$, $p < .001$, partial $\eta^2 = .27$. A 3 (Condition) $\times$ 2 (Time) analysis on each topic separately confirmed the observed difference between experimental conditions and control condition on the social topic, $F(2, 53) = 22.2$, $p < .001$, partial $\eta^2 = .46$, and on the science topic, $F(2, 53) = 3.95$, $p = .007$, partial

$\eta^2 = .13$. No significant difference between the two experimental conditions was observed for either topic.

Argumentation Skill at Initial and Final Assessment

Assessment of the quality of argumentation is based on the coding scheme described under "Coding Electronic Discourse Strategies." Analysis focused on those categories that accounted for greater than 5% of utterances, averaged across dialogues, at both initial and final assessment. These categories are Clarify, Counter-A (Counter-Alternative, consisting of disagreement together with proposal of an alternate argument), and Counter-C (Counter-Critique, consisting of disagreement accompanied by a critique of the opponent's argument; see Appendix F for examples). All other categories accounted for 5% or less of utterances at the initial and/or the final assessment. Counter-C, Counter-A, and Clarify are the argumentation strategies that in previous research have been found to either decrease (Clarify) or increase (the two types of Counterarguments) with practice (Felton, 2004; Kuhn, Goh, et al., 2008; Kuhn & Udell, 2003), as arguers begin to recognize the relevance of and accord more attention to their opponent's statements. Given the differences in number of utterances across time and conditions, percentages of usage were calculated for each participant, rather than frequencies. An arcsine transformation was used to normalize these proportions. To test the effect of conditions, a $3 \times 2 \times 2$ (Condition $\times$ Topic $\times$ Time) repeated-measures ANOVA was performed—with Topic and Time as the within-subject variables—as was a 3×2 (Condition $\times$ Time) repeated-measures ANOVA for each topic separately.

Use of Counterarguments on Intervention and Nonintervention Topic

In analyzing changes in counterargument usage across conditions, two indicators were employed. The first is the proportion of utterances that were coded as counterarguments, including both the more accomplished Counter-C strategy—which seeks to directly weaken the force of the opponent's preceding argument (see example in Appendix F)—and the less accomplished Counter-A strategy—which does not directly address the opponent's preceding argument but proposes an alternative argument (see example in Appendix F). The second is the proportion of utterances that were coded as Counter-Cs.

Overall counterarguments. An analysis of overall counterargument usage revealed a three-way interaction, $F(2, 53) = 11.08$, $p < .001$, partial $\eta^2 = .30$, for Condition $\times$ Time $\times$ Topic. Overall, participants at the initial assessment showed greater usage of counterargument strategies on the social topic than the science topic (suggesting that the social topic is more facilitative of counterargument). By the end of the intervention, however, participants in both experimental conditions increased their counterargument usage, with the SCI participants exhibiting equivalent achievement on both topics.

A two-way analysis on the social topic showed that there is a significant Time $\times$ Condition interaction, $F(2, 53) = 44.58$, $p < .001$, partial $\eta^2 = .63$. As shown in Figure 1, the two experimental conditions were equally effective in raising overall counterargument usage on the social topic, whereas the control condition was not effective, $F(2, 53) = 10.26$, $p < .001$, partial $\eta^2 = .28$. Of particular interest is the fact that participants in the SCI were able to show transfer of their counterargument skill to the social topic—in fact to the same extent as that shown by participants in the SOC, for whom it was their intervention topic.

A two-way analysis for the science topic also showed a significant Time $\times$ Condition interaction, $F(2, 53) = 51.73$, $p < .001$, partial $\eta^2 = .66$. Although the two experimental conditions were effective in raising overall counterarguments, compared with the control condition, $F(2, 53) = 36.75$, $p < .001$, partial $\eta^2 = .58$, the magnitude of their effectiveness was different,

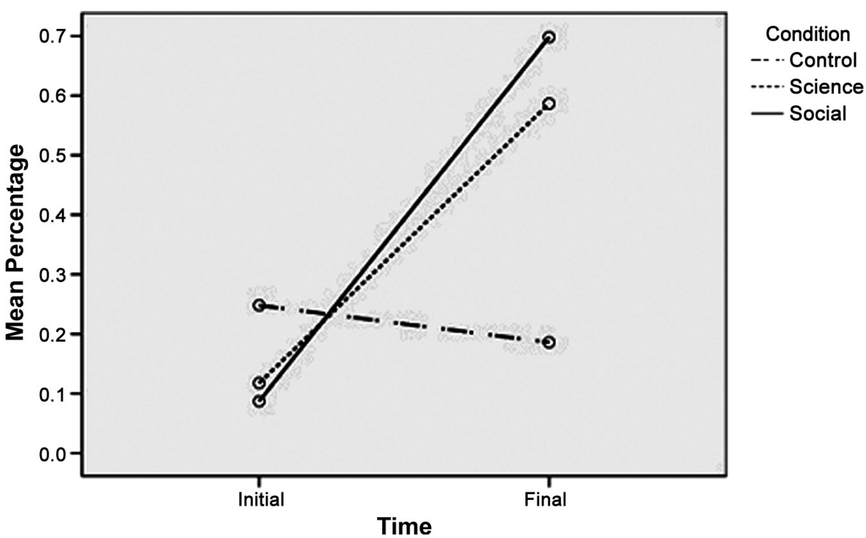


FIGURE 1 Percentage of utterances that were coded as overall counterargument on the social topic (HS) by condition.

$p < .001$. As seen in Figure 2, the SCI was more effective in raising overall counterargument usage ($M = 56.1\%$, $SD = 15.18$) on the science topic than the SOC ($M = 32.92\%$, $SD = 20.4$). However, Bonferroni post-hoc analysis revealed that the SOC produced significantly more counterarguments compared with the control condition ($M = 2.42\%$, $SD = 7.39$), demonstrating the ability of the SOC participants to transfer their counterargument skills to the nonintervention science topic.

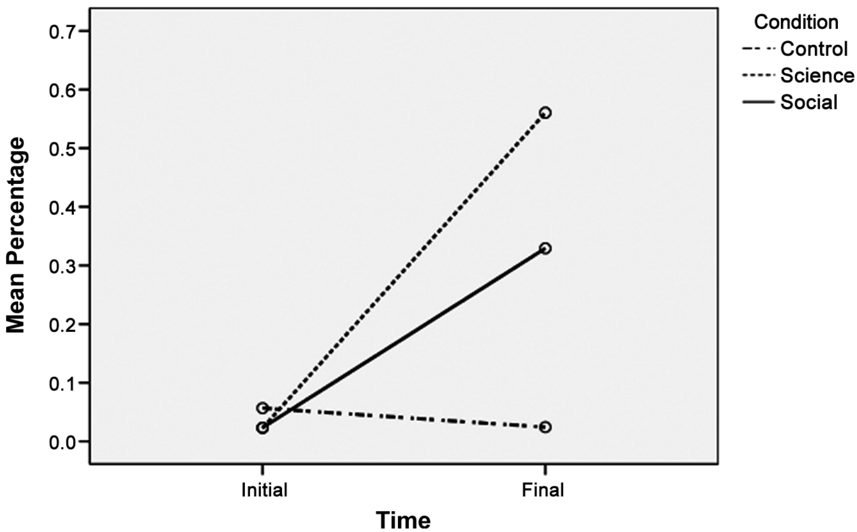


FIGURE 2 Percentage of utterances that were coded as overall counterargument on the science topic (DE) by condition.

Counter-C. A 3 × 2 × 2 (Condition × Time × Topic) repeated-measures ANOVA for Counter-C revealed a three-way interaction, $F(2, 53) = 4.59$, $p = .015$, partial $\eta^2 = .15$. Again, at the initial assessment, participants showed greater usage of Counter-C on the social topic. A separate two-way (Condition × Time) repeated-measures ANOVA for the social topic showed a significant Time × Condition interaction, $F(2, 53) = 24.81$, $p < .001$, partial $\eta^2 = .48$. The two experimental conditions were comparably effective in raising Counter-C usage, compared with the control condition, $F(2, 53) = 4.07$, $p = .023$, partial $\eta^2 = .13$. As seen in Figure 3, participants in the SOC increased from 7.03% ($SD = 10.46$) to 45.72% ($SD = 16.42$), and participants in the SCI increased from 9.09% (13.21) to 43.76% (15.35), whereas control participants showed no improvement.

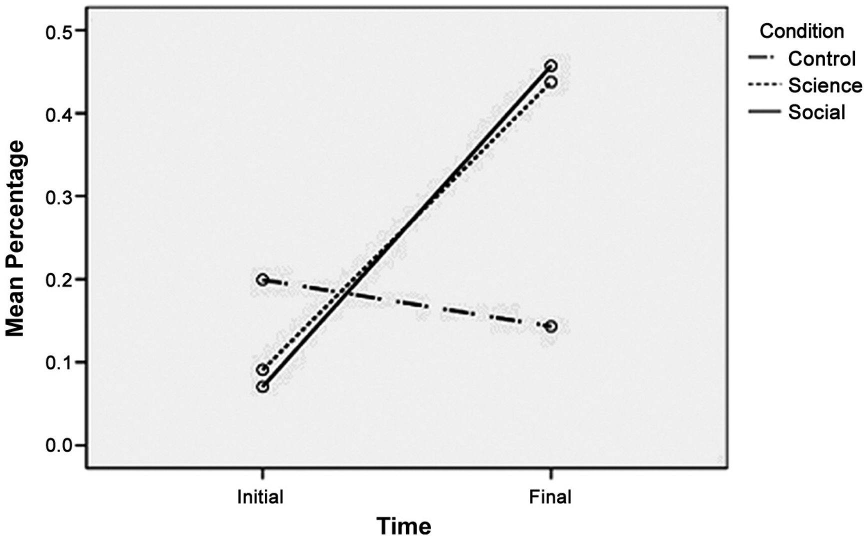


FIGURE 3 Percentage of utterances that were coded as Counter-C on the social topic (HS) by condition.

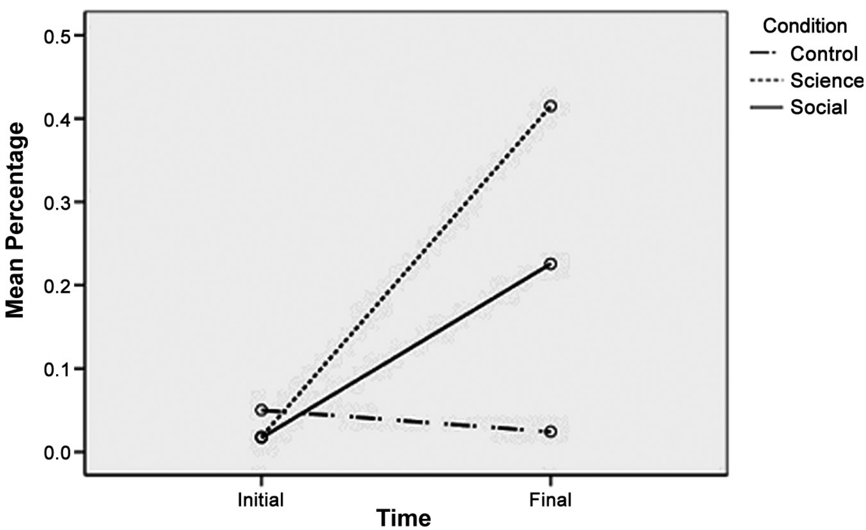


FIGURE 4 Percentage of utterances that were coded as Counter-C on the science topic (DE) by condition.

A two-way analysis for the science topic showed a significant Time × Condition interaction, $F(2, 53) = 28.03$, $p < .001$, partial $\eta^2 = .51$. As in the analysis of overall counterarguments, only participants in the experimental conditions showed an increase in Counter-C usage, $F(2, 53) = 20.18$, $p < .001$, partial $\eta^2 = .43$. However, as shown in Figure 4, the increase exhibited by participants in the SCI, from 1.83% (SD = 4.97) to 41.48% ($SD = 18.01$), was greater than the one exhibited by participants in the SOC, from 1.71% (5.26) to 22.56% (18.36). Those in the SOC nevertheless showed greater improvement in Counter-C usage in the science topic compared with the control group (Bonferroni post-hoc test, $p = .019$), demonstrating their transfer of Counter-C skill to the nonintervention topic.

Participants in both experimental conditions exhibited some transfer of their gains in Counter-C usage to the nonintervention topic. However, those in the SCI were able to transfer their Counter-C skills to the nonintervention topic to the same level that these skills were mastered in the intervention topic.

Individual patterns of change. In addition to analyses of group trends, equally informative are analyses of changes at the individual level. The percentage of participants who produced at least three counterarguments or Counter-Cs was examined. The criterion of "at least three" ensures that production of counterargument was not a random incident, but the result of significant mastery of the skill. Before examining change, I looked for evidence at the individual level to confirm the group pattern suggesting that the social topic is in general more facilitative of counterargument. Individual-level analysis further supported this finding: At initial assessment, no participants exhibited adequate mastery of Counter-C or counterargument usage in the science topic, whereas 18%—10 of 56—produced at least three Counter-Cs and 29%—16 of 56—produced at least three counterarguments on the social topic at the initial assessment (Fisher's Exact Test, $p < .001$).

Turning now to change, Table 1 presents the percentages (and numbers) of participants who made at least three Counter-Cs and counterarguments at initial and final assessment on both the intervention and nonintervention topics. Whereas both experimental conditions showed increased usage of counterargument strategies on both topics, none of the control condition participants produced more than three Counter-Cs or counterarguments on either the social or the science topics at the final assessment.

Finally, the discrepancy between SOC and SCI with respect to ability to transfer counterargument skill to a different domain was still distinctive. Only about half—11 of 19—of the SOC participants were able to produce at least three Counter-Cs on the science topic (transfer topic), whereas almost all—18 of 20—of the SCI participants did so on the social topic (transfer topic) at the final assessment.

TABLE 1
Initial and Final Percentages (and Number) of Participants Who Produced at Least Three Counter-Cs and at Least Three Counterarguments by Topic and Condition

		Counter-Cs		*Counterarguments*	
Topic	*Condition*	*Initial assessment*	*Final assessment*	*Initial assessment*	*Final assessment*
Social Topic	Social ($N = 19$)	21% (4)	95% (18)*	21% (4)	100% (19)**
	Science ($N = 20$)	5% (1)	90% (18)*	10% (2)	100% (20)**
	Control ($N = 17$)	29% (5)	0% (0)	59% (10)	0% (0)***
Science Topic	Social ($N = 19$)	0% (0)	58% (11)*	0% (0)	74% (14)**
	Science ($N = 20$)	0% (0)	90% (18)**	0% (0)	100% (20)**
	Control ($N = 17$)	0% (0)	0% (0)	0% (0)	0% (0)

*$p = .001$, McNemar test. **$p < .001$, McNemar test. ***$p < .002$, McNemar test.

Use of exposition (clarify) on the intervention and nonintervention topic. The expectation of a decline in the proportion of Clarify utterances as participants devoted more attention to counterargument was confirmed. A three-way ANOVA for Clarify revealed a Time × Condition significant interaction, $F(2, 53) = 6.96$, $p = .002$, partial $\eta^2 = .21$. These results were also supported by a two-way analysis for each topic separately. The two experimental conditions were equally effective in decreasing the proportion of utterances devoted to exposition of own position on both topics. A two-way analysis showed a significant Time × Condition interaction for both the social topic, $F(2, 53) = 5.50$, $p = .007$, partial $\eta^2 = .17$, and the science topic, $F(2, 53) = 3.84$, $p = .028$, partial $\eta^2 = .13$. SOC participants decreased from 30.26% ($SD = 18.97$) to 9.40% ($SD = 9.65$) on the social topic and from 30.26% (18.97) to 9.40% (9.65) on the science topic. Similarly, SCI participants decreased from 37.64% ($SD = 21.24$) to 14.75% (SD = 9.93) on the social topic and from 40.98% (21.46) to 12.96% (11.94) on the science topic. In contrast, control participants showed almost no change, exhibiting 38.83% (18.29) at initial assessment and 39.79% (17.98) at the final assessment on the social topic and 44.03% (23.98) at initial assessment and 42.45% (27.27) at the final assessment on the science topic.

Use of rebuttal on the intervention and nonintervention topic. A further important aspect of argument skill is the extent to which arguers are able to maintain focus and consistency in intent and execution to an extent that enables them to sustain the critique of one another's arguments. To examine this skill, both the frequency of rebuttals and the length of rebuttal strings were observed. Rebuttal is defined as a Counter-C immediately following a counterargument by the opposing partner (Felton & Kuhn, 2001; Kuhn & Udell, 2003). When the opponent critiques one's argument, through either a Counter-A or a Counter-C, the subject rebuts the opponent's critique by taking back the force of his or her own argument. Rebuttals entail a sequence of strategies involving both opponents—a rebuttal cannot be made in the absence of a counterargument by the opponent. Therefore, a different form of analysis is required than that employed for the counterargument strategies. Because participants' failure to make a rebuttal could be due either to lack of ability or lack of opportunity, to exclude the latter possibility, analysis of rebuttal included only those participants who had an opportunity to make a rebuttal.

On the social topic, only half of the experimental condition participants—17 of 39—had an opportunity to make a rebuttal at initial assessment, and of those who had an opportunity only half of them did so—9 of 17. A significantly higher proportion of control condition participants made a rebuttal—13 of 14—at initial assessment on the social topic ($a = .018$, Fisher-Irwin test). Most of the control condition participants had an opportunity to make a rebuttal and almost all of them did so. At the final assessment, however, the proportion of experimental condition participants who made a rebuttal—38 of 39—was significantly higher than the corresponding proportion of the control condition participants—2 of 8 ($a = .018$, Fisher-Irwin test). No significant difference was observed between the two experimental conditions on the social topic.

On the science topic, only a few participants had an opportunity to make a rebuttal at initial assessment—11 of 57—across all conditions, and only a few of these actually made a rebuttal—3 of 11. Yet, at the final assessment, the proportion of experimental condition participants who produced a rebuttal—33 of 37—was significantly higher than the proportion of control condition participants who did so—0 of 3 ($a < .001$, Fisher-Irwin test). In addition, a difference was observed in performance across the two experimental conditions. Although all of the SCI participants who had an opportunity to make a rebuttal did so—20 of 20—not all of the SOC participants who had an opportunity to make a Rebuttal did so—13 of 17 ($a = .036$, Fisher-Irwin test). This finding of the differential performance of the SOC and SCI participants on the science topic is consistent with the differences across conditions observed in overall counterargument and Counter-C usage analysis.

Finally, the length of rebuttal strings was examined. A length of 1 represents a sequence of assertion–counterargument–counterargument (rebuttal); a length of 2 represents a sequence consisting of assertion–counterargument–counterargument (rebuttal)–counterargument (rebuttal; see example in Appendix F).

Results showed that experimental condition participants who had the opportunity to make rebuttals increased the length of rebuttal sequences from initial to final assessment on both topics, whereas control condition participants who had rebuttal opportunity showed no increase from initial to final assessment. Among SOC participants having rebuttal opportunity, mean length of rebuttal increased from 2.5 to 3 on the social topic and from 1 to 2.25 on the science topic. Among the corresponding group of SCI participants, length of Rebuttal increased from 1 to 2.49 on the social topic and from 0 to 2.71 on the science topic. Due to the reduced sample size, however, a statistical analysis was not conducted.

Belief Revision

To examine whether argumentation taking place during the intervention affected participants' positions, their responses on the opinion scale—administered at initial and final assessment—were used. On the intervention topic, none of the participants in the SOC and only one of the participants in the SCI changed their position (the latter changed her position to "undecided"). Because argument skill was the focus of interest here, belief revision was not examined further.

Epistemological Levels

Table 2 presents by condition the number of participants who from initial to final assessment either relinquished an absolutist position (Table 2a) or attained an evaluativist position (Table 2b) in social and science domains. As seen in Table 2, evaluativist positions became more prevalent at the final

TABLE 2a
Number of Participants Showing an Absolutist Position at Initial and Final Assessment by Topic and Condition

Topic	*Condition*	*Initial assessment*	*Final assessment*
Social	Social ($N = 19$)	12 (63%)	2 (10.50%)*
	Science ($N = 20$)	14 (73%)	7 (36.80%)**
	Control ($N = 17$)	3 (17.64%)	11 (64.70%)****
Science	Social ($N = 19$)	7 (36%)	3 (15.80%)
	Science ($N = 20$)	9 (45%)	3 (15%)***
	Control ($N = 17$)	4 (23.50%)	3 (17.60%)

*Significant change, $p = .002$, McNemar test. **Significant change, $p = .016$, McNemar test. ***Significant change, $p = .031$, McNemar test. ****Significant change, $p = .008$, McNemar test.

TABLE 2b
Number of Participants Showing an Evaluativist Position at Initial and Final Assessment by Topic and Condition

Topic	*Condition*	*Initial assessment*	*Final assessment*
Social	Social ($N = 19$)	2 (10.50%)	8 (42%)*
	Science ($N = 20$)	3 (15%)	8 (40%)
	Control ($N = 17$)	5 (29.40%)	2 (11%)
Science	Social ($N = 19$)	3 (15.80%)	7 (36.82%)
	Science ($N = 20$)	5 (38.50%)	13 (65%)**
	Control ($N = 17$)	8 (47%)	9 (50%)

*Significant change, $p = .031$, McNemar test. **Significant change, $p = .008$, McNemar test.

assessment, compared with the initial assessment, but only in the domain of the intervention condition. That is, participants in the SCI showed significant advance in the science domain only, whereas participants in the SOC showed significant advance in the social domain only, and participants in the control condition showed no advance. Note, however, that for the social topic, the performance of participants in the SCI almost equaled that of participants in the SOC, whereas the reverse was not true. The SOC did not effect comparable progress in the science domain. It was in the science domain that the greatest change was observed, with participants becoming increasingly likely to accept the idea of multiple positions amenable to evaluation, but only among participants whose intervention was in the science domain.

DISCUSSION

The present study shows that argumentation skills in scientific domains are amenable to development as are skills in social domains, using an electronic dialogue method centered on engagement, practice, and reflection. This method proved also to be successful in producing transfer of argument skills across domains in both directions—from a science to social topic and from a social to science topic, a result critical in establishing the generality of such skills across different content. However, a difference in the magnitude of transfer was observed. The science condition did the better job of increasing levels of counterargument on the science topic, while the two conditions were equally effective in increasing levels of these strategies on the social topic. Discussion begins with the results for the domain in which participants engaged in the intervention and then proceeds to the issue of transfer of this skill across domains.

Development of Argument Skills Within a Domain

The intervention proved effective in developing participants' argumentation skills in the domain in which it was carried out. Participants exhibited an increased frequency of usage of advanced (counterargument and rebuttal) argument strategies and decreased frequency of less advanced (exposition) strategies within the context of their intervention topic. Although initially only a few participants exhibited the Counter-C strategy, by the end of the intervention all did so. Regarding the advanced strategy of rebuttal, of the participants who had an opportunity to make a rebuttal, only a few made a rebuttal at initial assessment, whereas all did so in the science condition and all but one in the social condition at the final assessment. In addition, participants exhibited increased proportion of usage of the more advanced Counter-C strategy in contrast to the less advanced Counter-A strategy. Social condition participants doubled the percentage of counterarguments that were Counter-C in the social domain, and science condition participants' percentage was four times greater at the final compared with initial assessment in the science domain.

In addition to changes in strategy usage, corresponding increases were observed in the more surface indicators of both number and length of utterances. The length of utterances produced at the final assessment was almost twice the length of utterances produced at initial assessment on the intervention topic for both social and science conditions. These findings on quantitative output are consistent with previous research, both in the social (Kuhn & Udell, 2003) and science domains (Mercer & Littleton, 2007; Naylon et al., 2007; Zohar & Nemet, 2002).

The present findings in the social domain are consistent with findings of previous cross-sectional (Felton & Kuhn, 2001) and experimental studies of developing argumentation skills in the social domain, using similar methods (Felton, 2004; Kuhn, Goh, et al., 2008; Kuhn & Udell, 2003; Udell, 2007) as well as methods based on similar theoretical principles (Anderson et al., 2001; Nussbaum, 2005; Nussbaum & Sinatra, 2003; Reznitskaya et al., 2001). Specifically, improvement by participants in the experimental conditions is comparable to the improvement observed by Kuhn and Udell (2003; from 5.30% to 30.60%) and more impressive compared with Kuhn, Goh, et al. (2008; from 4.50% to 21.37%).

A significant contribution of the present study is its documentation of the efficiency of discourse-based methods in developing argumentation skills in the science domain. Although the central role of argument to science and science education has been widely endorsed by science educators (Driver, Newton, & Osborne, 2000; Erduran, Simon, & Osborne, 2004; Kelly, Regev, & Prothero, 2008; Lehrer, Schauble, & Petrosino, 2001), developing these skills in science students has proven challenging (Osborne et al., 2004) and understanding of mechanisms of development is at best incomplete. The present findings show that forms of engagement and practice that have been shown to support development of argumentation skills in the social domain can also support this development in the science domain. This result establishes that students' limited argument skills in the science domain reported in several studies (Driver et al.; Solomon, 1992) are not due to constraints imposed by the nature of the science domain itself.

Control condition performance, showing no improvement in either the social or the science domain, establishes that the limited opportunities offered by the regular curriculum to practice argumentation or merely the passage of time are not sufficient to advance students' argumentation skills. Also, it should be noted that control condition participants' performance at initial assessment was comparable to that of participants in the experimental condition, and a few control condition participants even showed a slight advantage over experimental condition participants at the outset, making the study's findings even stronger. However, since intact classes, rather than individual students, were randomly assigned to control and experimental conditions, comparisons between the two groups should be interpreted with caution. The decline in performance of control condition participants from initial to final assessment suggests the critical role of participants' interest and motivation in developing the kinds of cognitive skills examined here. For control condition participants, final assessment was a mere repetition of the activity they engaged in 2 months earlier, and they may have seen little value in repeating it. Among the experimental condition participants, in contrast, the goal-based nature of the activities proved effective in maintaining their interest and involvement throughout the intervention.

Development of Argument Skills Across Domains

Regarding the critical question of the transfer of argumentation skills across domains, the present study's results show that transfer does occur across topics in different domains. Post-intervention performance on the nonintervention topic was superior to control-condition performance in both conditions. How was this transfer achieved? One mechanism that may contribute to the transfer of argument skills across domains is the development of meta-level awareness and understanding of the objectives of argument and in particular of the relevance of the other person's position. It is possible that this developing meta-level understanding supports the execution of argument skills at the procedural level across domains of application. This possibility is further supported by the present study's findings of development in levels of epistemological understanding in both domains, although here some evidence of asymmetry was also found across domains, parallel to that in the main findings.

Although participants in both conditions exhibited transfer, a difference in the magnitude of transfer was observed, with those in the science condition showing transfer to a greater extent than those in the social condition—in fact, their counterargument skills in the nonintervention topic reached the same level that they did in the intervention topic.

A possible explanation for the condition difference observed is that the social topic is more facilitative of counterargument, a difference documented at the initial assessment. Participants overall exhibited greater counterargument usage on the social topic than they did on the science topic, a finding consistent with previous research (Osborne et al., 2004). Further research, using multiple topics in each domain, is required to better illuminate our understanding of differences between science and social domains. The fact that the present intervention included only a single topic in each domain indicates that further studies are needed with a wider range of topics in each domain before strong conclusions can be drawn regarding differences between science and social domains (Kuhn, in press; Kuhn, Iordanou, et al., 2008). The encouraging message, however, that the present study carries is that argumentation skills in the science domain are amenable to development. This was true in the present study even to the extent of overcoming and erasing initial performance difference across domains.

Implications for Science Education

Beyond their implications for understanding the nature and development of argument skills, the present findings have implications for science education. Argument is central to the process by which science advances (Duschl & Osborne, 2002; Kuhn, 1993; Lehrer et al., 2001), and students need to become both aware of its relevance and skilled in its execution. Observations of science lessons suggest that activities that involve argumentation are scant (Driver et al., 2000; Naylon et al., 2007; Solomon, 1992). Even in activities involving experimentation, students are guided to "discover" the one right answer that unproblematically evolves from their experimentation (Lehrer et al., 2001), reinforcing naïve epistemological conceptions of science as accumulating fact. Studies of students' argument skills in science contexts report these skills to be weak at best, even at the college level (Kelly, Druker, & Chen, 1998; Kelly & Takao, 2002; Kolstø et al., 2006; Maloney, 2007; Maloney & Simon, 2006; Sadler, 2004).

In the present work, such weaknesses are addressed by exploring elementary skills of dialogic argumentation as a path toward the development of competence in constructing and evaluating authentic scientific arguments. The virtues of such a path are several. Discourse makes thinking visible to students (Duschl, 2008). Furthermore, as Graff (2003) emphasizes, by concretizing an alternative position (without which argument has little point), dialogic argument helps students to appreciate the purpose and point of individual, expository argument. This may be a necessary first step toward enabling students to integrate and coordinate evidence with claims in order to construct strong expository arguments in content-rich scientific domains.

The present findings support the view that attention to the development of argumentation skill (and its associated epistemological understanding) within science domains is warranted. Argument skill in the science domain is amenable to the same development as has been shown for argument skill in the social domain, but specific engagement and practice within the science domain may be required for optimum development of such skill. The policy recommendation supported by the present findings—engagement and practice in argumentation within the context of meaningful science topics—is consistent with the educational objective of fostering students' competence to assume roles in meaningful scientific discourse, rather than become merely consumers of scientific facts.

ACKNOWLEDGEMENT

This research was submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Columbia University, under the direction of Deanna Kuhn.

REFERENCES

Anderson, R., Nguyen-Jahiel, K., McNurlen, B., Archodidou, A., Kim, S., Reznitskaya, A., ... Gilbert, L. (2001). The snowball phenomenon: Spread of ways of talking and ways of thinking across groups of children. *Cognition & Instruction*, *19*(1), 1–46.

Andriessen, J. (2006). Arguing to learn. In K. R. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 443–459). New York, NY: Cambridge University Press.

Baron, J. (1995). Myside bias in thinking about abortion. *Thinking and Reasoning*, *1*(3), 221–235.

Bell, P., & Linn, M. (2000). Scientific arguments as learning artifacts: Designing for learning from the Web with KIE. *International Journal of Science Education*, *2*(8), 797–817.

Bereiter, C. (2002). *Education and mind in the knowledge age*. Mahwah, NJ: Erlbaum.

Billig, M. (1987). *Arguing and thinking: A rhetorical approach to social psychology*. Cambridge, United Kingdom: Cambridge University Press.

Clark, D. B., & Sampson, V. D. (2007). Personally seeded discussions to scaffold online argumentation. *International Journal of Science Education*, *29*(3), 253–277.

Clark, D. B., Stegmann, K., Weinberger, A., Menekse, M., & Erkens, G. (2008). Technology-enhanced learning environments to support students' argumentation. In S. Erduran &

M. P. Jiménez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 217–243). Dordrecht, The Netherlands: Springer.

Dean, D., & Kuhn, D. (2007). Direct instruction versus discovery: The long view. *Science Education*, *91*, 384–397.

De Fuccio, M., Kuhn, D., Udell, W., & Callender, K. (2009). Developing argument skills in severely disadvantaged adolescent males in a residential correctional setting. *Applied Developmental Science*, *13*(1), 30–41.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, *84*(3), 287–312.

Duschl, R. A. (2008). Quality argumentation and epistemic criteria. In S. Erduran & M. P. Jiménez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 137–157). Dordrecht, The Netherlands: Springer.

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

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

Felton, M. (2004). The development of discourse strategies in adolescent argumentation. *Cognitive Development*, *19*, 35–52.

Felton, M., & Kuhn, D. (2001). The development of argumentative discourse skill. *Discourse Processes*, *32*, 135–153.

Graff, G. (2003). *Clueless in academe: How schooling obscures the life of the mind*. New Haven, CT: Yale University Press.

Greene, A. J., Azevedo, R., & Torney-Purta, J. (2008). Modeling epistemic and ontological cognition: Philosophical perspectives and methodological directions. *Educational Psychologist*, *43*(3), 142–160.

Hofer, B., & Pintrich, P. (Eds.). (2002). *Epistemology: The psychology of beliefs about knowledge and knowing*. Mahwah, NJ: Erlbaum.

Jiménez-Aleixandre, M. P., & Erduran, S. (2008). *Argumentation in science education: An overview*. In S. Erduran & M. P. Jiménez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 137–157). Dordrecht, The Netherlands: Springer.

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.

Kelly, G. J., Regev, J., & Prothero, W. (2008). Analysis of lines of reasoning in written argumentation. In S. Erduran & M. P. Jiménez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 137–157). Dordrecht, The Netherlands: Springer.

Kelly, G. J., & Takao, A. (2002). Epistemic levels in argument: An analysis of university oceanography students' use of evidence in writing. *Science Education*, *86*(3), 314–342.

Klaczynski, P. A. (2000). Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: A two-process approach to adolescent cognition. *Child Development*, *71*, 1347–1366.

Klaczynski, P. A., & Gordon, D. H. (1996). Self-serving influences on adolescents' evaluations of belief-relevant evidence. *Journal of Experimental Child Psychology*, *62*(3), 317–339.

Klaczynski, P. A., & Lavallee, K. L. (2005). Domain-specific identity, epistemic relation, and intellectual ability as predictors of belief-biased reasoning: A dual-process perspective. *Journal of Experimental Child Psychology*, *92*, 1–24.

Klahr, D. (2000). *Exploring science: The cognition and development of discovery processes*. Cambridge, MA: MIT Press.

Kolstø, S. D., Bungum, B., Arnesen, E., Isnes, A., Kristensen, T., Mathiassen, K., . . . Ulvik, M. (2006). Science students' critical examination of scientific information related to socioscientific issues. *Science Education*, *90*(4), 632–655.

Kuhn, D. (1991). *The skills of argument*. New York, NY: Cambridge University Press.

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

Kuhn, D. (2005). *Education for thinking*. Cambridge, MA: Harvard University Press.

Kuhn, D. (2009). The importance of learning about knowing. *Child Development Perspectives*, *3*(2), 112–117.

Kuhn, D. (in press). Reasoning. In P. Zelazo (Ed.), *Oxford handbook of developmental psychology*. New York, NY: Oxford University Press.

Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development*, *15*, 309–328.

Kuhn, D., & Dean, D. (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science*, *16*, 866–870.

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

Kuhn, D., Goh, W., Iordanou, K., & Shaenfield, D. (2008). Arguing on the computer: A microgenetic study of developing argument skills in a computer-supported environment. *Child Development*, *79*(5), 1310–1328.

Kuhn, D., Iordanou, K., Pease, M., & Wirkala, C. (2008). Beyond control of variables: What needs to develop to achieve skilled scientific thinking? *Cognitive Development*, *23*, 435–451.

Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentative reasoning. *Cognition and Instruction*, *15*, 287–315.

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

Lehrer, R., & Schauble, L. (2006). Scientific thinking and science literacy. In W. Damon, R. Lerner, K. A. Renninger, & I. E. Sigel (Eds.), *Handbook of child psychology: Vol. 4. Child psychology in practice* (6th ed.). Hoboken, NJ: John Wiley & Sons.

Lehrer, R., Schauble, L., & Petrosino, A. J. (2001). Reconsidering the role of experiment in science education. In K. Crowley, C. Schunn, & T. Okada (Eds.), *Designing for science: Implications from everyday, classroom, and professional settings* (pp. 251–277). Mahwah, NJ: Erlbaum.

Maloney, J. F. (2007). Children's roles and use of evidence in science: An analysis of decision-making in small groups. *British Educational Research Journal*, *33*(3), 371–401.

Maloney, J., & Simon, S. (2006). Mapping children's discussions of evidence in science to assess collaboration and argumentation. *International Journal of Science Education*, *28*(15), 1817–1841.

Mason, L. (1998). Sharing cognition to construct scientific knowledge in school context: The role of oral and written discourse. *Instructional Science*, *26*(5), 359–389.

Mason, L. (2003). Personal epistemologies and intentional conceptual change. In G. M. Sinatra & P. R. Pintrich (Eds.), *Intentional conceptual change* (pp. 199–236). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Mason, L., & Scirica, F. (2006). Prediction of students' argumentation skills about controversial topics by epistemological understanding. *Learning and Instruction*, *16*, 492–509.

Means, M., & Voss, J. (1996). Who reasons well? Two studies of informal reasoning among students of different grade, ability, and knowledge levels. *Cognition & Instruction*, *14*, 139–178.

Mercer, N., & Littleton, K. (2007). *Dialogue and development of children's thinking: A sociocultural approach.* London, United Kingdom; New York, NY: Routledge.
Muis, K., Bendixen, L., & Haerle, F. (2006). Domain generality and domain specificity in personal epistemology research: Philosophical and empirical reflections in the development of a theoretical framework. *Educational Psychology Review*, *18*, 3–54.
Naylon, S., Keogh, B., & Downing, B. (2007). Argumentation and primary science. *Research in Science Education*, *37*, 17–39.
Nussbaum, E. M. (2005). The effects of goal instructions and need for cognition on interactive argumentation. *Contemporary Educational Psychology*, *30*, 286–313.
Nussbaum, E. M., & Sinatra, G. M. (2003). Argument and conceptual engagement. *Contemporary Educational Psychology*, *28*, 384–395.
Oaksford, M., Chater, N., & Hahn, U. (2008). Human reasoning and argumentation: The probabilistic approach. In J. Adler & L. Rips (Eds.), *Reasoning: Studies of human inference and its foundations*. New York, NY: Cambridge University Press.
Osborne, G., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, *41*(10), 994–1020.
Perkins, D. N. (1985). Postprimary education has little impact on informal reasoning. *Journal of Educational Psychology*, *77*, 562–571.
Reznitskaya, A., Anderson, R., McNurlen, B., Nguyen-Jahiel, K., Archodidou, A., & Kim, S. (2001). Influence of oral discussion on written argument. *Discourse Processes*, *32*, 155–175.
Sadler, T. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching*, *41*, 513–536.
Solomon, J. (1992). The classroom discussion of science-based social issues presented on television: Knowledge, attitude, and values. *International Journal of Science Education*, *14*(4), 431–444.
Stanovich, K. E., Toplak, M. E., & West, R. F. (2008). The development of rational thought: A taxonomy of heuristics and biases. *Advances in Child Development and Behavior*, *36*, 251–285.
Stanovich, K. E., & West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Thinking & Reasoning*, *13*(3), 225–247.
Strand-Cary, M., & Klahr, D. (2008). Developing elementary science skills: Instructional effectiveness and path independence. *Cognitive Development*, *23*, 488–511.
Udell, W. (2007). Enhancing adolescent girls' argument skills in reasoning about personal and non-personal decisions. *Cognitive Development*, *22*(3), 341–352.
Wagner, T. (2008). *The global achievement gap*. New York, NY: Basic Books.
Walton, D. N. (1989). Dialogue theory for critical thinking. *Argumentation*, *3*, 169–184.
Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review*, *20*, 99–149.
Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*, 172–223.
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.

APPENDIX A

Argument Skill Goals (From Kuhn & Udell, 2003)

GENERATING REASONS
Goals: Reasons underlie opinions.
Different reasons ->same opinion

ELABORATING REASONS
Goal: Good reasons support opinions.

SUPPORTING REASONS WITH EVIDENCE
Goal: Evidence can strengthen reasons.

EVALUATING REASONS
Goal: Some reasons are better than others.

DEVELOPING REASONS INTO AN ARGUMENT
Goal: Reasons connect to one another and are building blocks of argument.

EXAMINING AND EVALUATING OPPOSING SIDE'S REASONS
Goal: Opponents have reasons too.

GENERATING COUNTERARGUMENTS TO OTHERS' REASONS
Goal: Counters to reasons can be rebutted.

GENERATING REBUTTALS TO OTHERS' COUNTERARGUMENTS
Goal: Counters to reasons can be rebutted.

CONTEMPLATING MIXED EVIDENCE
Goal: Evidence can be used to support different claims.

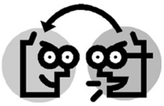
CONDUCTING AND EVALUATING TWO-SIDED ARGUMENTS
Goal: Some arguments are stronger than others.

APPENDIX B

'Other Argument' Reflection Sheet

Team members____
Date ________

Let's think . . . Starting with the other side's argument



*One of the **other side's** MAIN ARGUMENTS was:*	*Our COUNTERARGUMENT **against** their argument was:*
____________________	____________________
____________________	____________________
____________________	____________________
____________________	____________________
____________________	____________________
____________________	____________________
____________________	____________________
____________________	____________________
____________________	____________________
____________________	____________________

How can this COUNTERARGUMENT be **improved?**
Is there a more effective counterargument?
____
____

APPENDIX C

'Own Argument' Reflection Sheet

Team members ____________________________________

Date ________

Let's think . . . Starting with our argument

*One of **our** MAIN ARGUMENTS was:*	***Their** COUNTERARGUMENT against our argument was:*	***Our** COMEBACK was:*

How can this COMEBACK be **improved?**
Is there a more effective comeback?

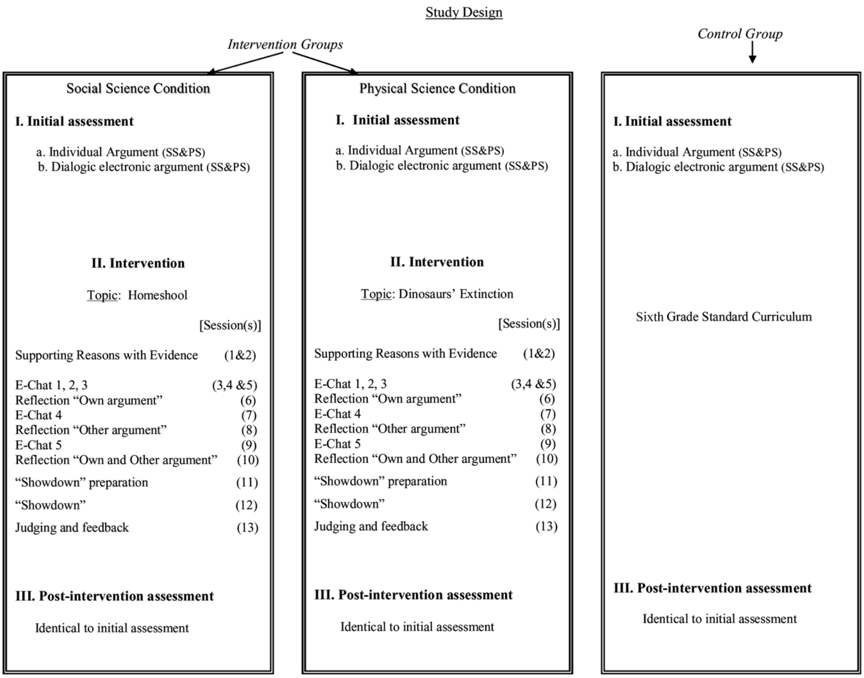
____

____

____

APPENDIX D

Outline of the Study Design



SS: Social Science topic

PS: Physical Science topic

At initial and final assessment, all participants—in all conditions—engaged in individual and dialogic arguments on both the social science topic and the physical science topic. In the intervention, the *Social Science Condition* engaged in a series of e-chats and reflective activities on the home-school topic. The *Physical Science Condition* engaged in the same activities but on the dinosaur extinction topic. The *Control Condition* engaged in the regular sixth-grade curriculum.

APPENDIX E

Summary of Utterance Types in the Analytic Scheme for Coding Argument Dialogue

Transactive questions

Agree-?	A question that asks whether the partner will accept or agree with the speaker's claim
Case-?	A request for the partner to take a position on a particular case or scenario
Clarify-?	A request for the partner to clarify his or her preceding utterance
Justify-?	A request for the partner to support his or her preceding claim with evidence or further argument
Meta-?	A question regarding the dialogue itself (vs. its content)
Position-?	A request for the partner to state his or her position on an issue
Question-?	A simple informational question that does not refer back to the partner's preceding utterance
Respond-?	A request for the partner to react to the speaker's utterance

Transactive statements

Add	An extension or elaboration of the partner's preceding utterance
Advance	An extension or elaboration that advances the partner's preceding argument
Agree	A statement of agreement with the partner's preceding utterance
Aside	A comment that does not extend or elaborate the partner's preceding utterance
Clarify	A statement or clarification of speaker's own argument in response to the partner's preceding utterance
Counter-A	A disagreement with the partner's preceding utterance, accompanied by an alternate argument
Counter-C	A disagreement with the partner's preceding utterance, accompanied by a critique of that argument

Disagree	A simple disagreement without further argument or elaboration
Dismiss	An assertion that the partner's immediately preceding utterance is irrelevant to the speaker's position
Interpret	A paraphrase of the partner's preceding utterance with or without further elaboration
Null	An unintelligible or off-task utterance
Refuse	An explicit refusal to respond to the partner's preceding question
Substantiate	An utterance offered in support of the partner's preceding utterance

Meta categories Utterances regarding the dialogue itself (vs. its content)

Meta-Argument	An utterance in reference to the content of an argument
Meta-Argumentation	An utterance in reference to the dialogic process
Meta-Directive	An utterance instructing the partner to do something
Meta-Comprehension	Utterance regarding the partners' understanding of one another
Meta-Task	Utterance in reference to execution of the task, including clarification of scenario, computer issues, spelling, etc.

Nontransactive statements

Continue	A continuation or elaboration of the speaker's own last utterance that ignores the partner's immediately preceding utterance
Unconnected	An utterance having no apparent connection to the preceding utterances of either partner or speaker

APPENDIX F

Example of Major Coding Categories From Participants' Electronic Dialogues

Example of Counter-C

Assertion
A: "I don't think it's possible for many volcanoes to have erupted at the same time."

Counter-C
B: "But one huge volcano could have erupted."

Example of Counter-A

Assertion
A: "Tsunami happened in one area and it couldn't affect mountain areas."

Counter-A
B: "They might have died during the Ice Age."

Example of Rebuttal Chain (Length 3)

Assertion
A: "Lava spreads out everywhere."

Counter-C
B: "Some dinosaurs could jump in the sea that is near to the island."

Counter-C (Rebuttal)
A: "Yes, but they will drown."

Counter-C (Rebuttal)
B: "Some dinosaurs may make it to a neighboring island."