Child Development, September/October 2008, Volume 79, Number 5, Pages 1310–1328

Arguing on the Computer: A Microgenetic Study of Developing Argument Skills in a Computer-Supported Environment

Deanna Kuhn, Wendy Goh, Kalypso Iordanou, and David Shaenfield
Columbia University

We report a study of a class of 28 sixth graders engaged in an extended computer-supported argumentive discourse activity. Participants collaborated with a same-side peer in arguing against successive pairs of peers on the opposing side of an issue. Meta-level awareness was facilitated by conducting the dialogs via instant messaging software, which made available a transcript of the dialog that was used in additional reflective activities. In the course of dialogs on 3 successive topics, participants showed significant gains in meta-level communications about the discourse, reflecting at least implicit understanding of its goals, as well as in the strategic moves that constituted the discourse. The latter advances remained evident when the social support of a same-side partner was withdrawn.

The central place of argument skills in the development of higher order thinking is increasingly being recognized. Indeed, the construction and evaluation of arguments constitute a major component of the College Learning Assessment (Hersh, 2005), designed to assess the cognitive development that occurs during the college years. Until recently, research on argument skills has been devoted almost entirely to individuals' written or verbal arguments in support of a claim. Studies have been consistent in revealing skills that are at best modest among children and adolescents and show little improvement thereafter (Brem & Rips, 2000; Glassner, Weinstock, & Neuman, 2005; Kuhn, 1991; Means & Voss, 1996; Perkins, 1985; Stanovich, 1999; Voss & Means, 1991; Weinstock, Neuman, & Tabak, 2004).

The work presented here is part of a growing body of research that examines the process of argumentation, or argumentive discourse, in older children and adolescents (Voss, 2001). As elaborated elsewhere (Graff, 2003; Kuhn, 2005), such discourse is of interest for its alleged role in facilitating the development of individual (nonsocial) argument skills. We agree with Graff's (2003) view that dialogic argument provides the "missing interlocutor" that lends purpose to and thereby facilitates development of individual argument skills.

Although competence in individual argument skills is clearly critical, most notably in academic contexts, the development of dialogic argumentation skills is important in its own right. What needs to develop in the domain of argumentive discourse skill and how does this development occur? The present work is focused on these questions, although we include an assessment of the extent to which gains observed in argumentive discourse affect performance on a measure of individual argument.

Dialogic argumentation has long been of broad interest to developmentalists, particularly those who regard social collaboration as central to cognitive development (Damon, 1984; Moshman, 2005; Smetana, Killen, & Turiel, 1991; Wainryb, Shaw, Laupa, & Smith, 2001). The focus of more recent, educationally oriented research on argumentive discourse has been on devising scaffolds (typically computer software) to support the argumentation process (Andriessen, 2006; Andriessen, Baker, & Suthers, 2004; Bell & Linn, 2000; Chinn, 2006; Clark & Sampson, 2005; Glassner & Schwarz, 2005; Nussbaum, 2005; Wiley & Voss, 1999; Zohar & Nemet, 2002). The work of Anderson and colleagues (Anderson et al., 2001; Reznitskaya et al., 2001) and, following them, Nussbaum (2003) is an exception in examining naturally occurring discourse in small groups of elementary-school-age children and seeking to identify the roots of more sophisticated argumentation skills as they emerge in such conversation and spread from one child to others.

Dialogic argument is a skill that has the advantage of early roots in everyday conversation, and yet, the demands of skilled argumentive discourse are considerable, with previous research indicating that few individuals attain proficiency during the second decade of life (Felton & Kuhn, 2001; Kuhn & Franklin, 2006; Kuhn & Udell, 2003; Udell, 2007). At the same time that one is processing and evaluating input from the conversational partner, one must be formulating

Correspondence concerning this article should be addressed to Deanna Kuhn, Box 119, Department of Human Development, Teachers College Columbia University, New York, NY 10027. Electronic mail may be sent to dk100@columbia.edu.

an effective response that meets discourse goals. According to Walton (1989), skilled argumentation has two goals. One is to secure commitments from the opponent that can be used to support one's own argument. The other is to undermine the opponent's position by identifying and challenging weaknesses in his or her argument. Both these goals, note, require attention to the opponent's claims.

In earlier work, we have found that young adolescents concentrate their efforts on exposition of their own claims to the neglect of attending to the opponent's claims and attempting to weaken their force (Felton, 2004; Felton & Kuhn, 2001; Kuhn & Udell, 2003; Udell, 2007). Yet, they are able to attend to the other's argument, we have found, and even generate an argument against it, when explicitly instructed to do so (Kuhn & Udell, 2007). The challenge, then, may be less one of executing the skill (of addressing the opposing position) than it is one of recognizing the need to do so. This recognition goes to the very heart of argument. If the opponent's argument for his or her position is not relevant, the process through which one may claim to attain victory over the opponent cannot be regarded as one of argument.

In the present work, we focus on how young adolescents develop better meta-level understanding of the goals of argument, as well as procedural skill in implementing these goals in discourse. Our prior work indicates that dense engagement over time in argumentive discourse with a series of peers does lead to advancements in procedural skill, and we seek to corroborate this finding in the present study. Our focus here, however, is on the development of meta-level understanding about argumentive discourse and its goals. Specifically, does this meta-level understanding also develop and support skill development at the procedural level, that is, in the strategies involved in the actual conduct of the argumentation?

Most generally, one must see a point to argument if one is to engage in it effectively (Kuhn, 1991, 2005). Specifically, we claim, one must understand the goal of argumentation as one of engaging one another's claims and undertaking to weaken them, as well as seeking acceptance of one's own claims—the two goals noted earlier identified by Walton (1989). Progress with respect to this meta-level development, we propose, is likely to support progress in the procedural aspect as well.

The form of dense engagement in peer discourse devised for this study was designed with the objective of supporting to the extent possible the meta-level, as well as strategic, development to which we have referred. Beyond dense engagement in discourse itself, which has been found effective with various age groups in our own previous studies (Felton, 2004; Felton & Kuhn, 2001; Kuhn, Shaw, & Felton, 1997; Kuhn & Udell, 2003; Udell, 2007) and those of several other researchers (Anderson et al., 2001; Chinn, 2006; Nussbaum, 2003; Nussbaum & Kardash, 2005; Nussbaum & Sinatra, 2003; Reznitskaya et al., 2001; Schwarz, Newman, & Biezuner, 2000), the general strategy we thought might be most effective is to heighten participants' awareness of the discourse. This meta-level awareness regarding the discourse seems a necessary condition to formulating, and ultimately implementing, goals with respect to it.

We employ three techniques designed to heighten and support this awareness. First, we ask participants to work in pairs. The two members of the pair hold the same position on an issue and they collaborate in arguments against a series of opposing pairs who hold the opposite position. Each communication to the opposing dyad is made only when the pair has agreed what it is to be. Having to deliberate regarding each argumentive move, we hypothesized, would not only generate more thoughtful contributions to the discourse but also make the collaborative pair more aware and reflective regarding this discourse.

The second technique, following on the study by Felton (2004), is to integrate an explicit reflective activity into the series of dialogs. In the present study, we do this by making available to participants some of their own dialog transcripts as the basis for this activity. Support for these design features comes from the study by Felton; although adolescents individually engaged in argument with an opposing peer, two peer advisors observed and later discussed possible improvements with their partners—a condition Felton found enhanced dialogic argument skills relative to a condition in which participants argued without peer advisors. In the present study, we engage peers in collaboratively planning and implementing the exchange itself, not only reflecting on it.

The third technique we employ here is to conduct these dialogs via computer-supported instant messaging (IM) software, a method we have not tried in previous work and the one that constitutes the major innovation of the present study. In addition to capitalizing on young teens' familiarity with the medium, this technique has the benefit of providing an immediately available and permanent record of the discourse, in striking contrast to the conditions of real-time verbal discourse, where the contents of each contribution to the dialog immediately disappear as soon as they are spoken.

The present study extends over an entire school year. We begin with assessments of individuals' argument skills, both in generating an individual

written argument in support of their position on a social issue and in engaging in an electronically conducted argument with a peer who holds an opposing view on the topic. These assessments are repeated at the end of the year, allowing us to assess improvement in both individual and dialogic argument skills. A focus of our work, however, is microgenetic analysis (Kuhn, 1995; Miller & Coyle, 1999; Siegler, 2006) of the period of intervention itself. During this time, we observe participants developing meta-level understanding of argumentation, as well as developing strategic skill in executing it, as each participant and a partner engage in electronic discourse with an opposing pair on a series of new topics.

Method

Participants

The 28 participants consisted of the entire sixth grade at a university-affiliated independent school in a large urban setting. The school has a unique population in that 50% of the school's slots are reserved for children of university faculty and high-level administrators, whereas the remaining 50% of slots are filled by children from the surrounding low- to middle-income community chosen by lottery and receiving full financial support. The school was in only its 2nd year of operation, the year that the work described here took place. Hence, students had come from a wide range of previous schools and were of diverse academic, as well as ethnic, racial, and socioeconomic, backgrounds. Roughly half were identifiable as racial or ethnic minorities (largely Hispanic and African American). All were 11 or 12 years of age; 16 were girls and 12 were boys.

Participants had become familiar with the school's innovative curriculum by the time our project began. The school features an integrated curriculum across all subjects and has a strong technology emphasis, with Smartboards in each classroom and students each allocated an individual laptop computer used daily in their schoolwork.

Initial Assessment

Capital punishment (CP) was the topic on which initial and final assessments of both individual and dialogic argument were based.

Individual argument. Each participant's initial position on CP was assessed using a 13-point Likert scale. The middle position on the scale read, ''I have mixed or undecided opinions about capital punishment.'' Each position to one side or the other of this middle point represents a position progressively more extreme in support of or against CP. The Point 1 position from the center on the ''support'' end of the scale, for example, read, ''I am somewhat in favor of CP but I'm not sure,'' whereas the end position read, ''I am totally in favor of CP. I will never change my opinion and I cannot imagine how someone could have a different opinion.'' Prior to asking participants to check a box that best represented their view, one of the adult coaches defined CP for the class (''administering the death penalty for serious crimes such as murder'') and explained the scale.

Completion of the scale instrument was followed by a written assessment instrument. After again eliciting the participant's position as for, against, or undecided, the following questions were posed (with space after each for the participant to respond):

1. Why did you choose this position? In other words, what are your reasons for your position?
2. What are some other possible reasons for this position? These reasons can be your own reasons or other people's reasons who have the same position as you.
3. Can you imagine someone having another position that is different from your own? What would it be?
4. What might be some reasons for having this different position?
5. Can you think of any other reasons to be for or against CP?

Dialogic argument. The 8 participants who indicated they were undecided on the scale gave reasons on both sides of the issue in the questionnaire. These 8 were grouped with the 6 participants who chose a scale position supportive of CP. The remaining 14 participants, who indicated they were opposed to CP on both the scale and the questionnaire, formed a second group. The week following the individual assessment, each of the participants from Group 1 was paired with a participant from Group 2, and each of these pairs was asked to participate in a dialog with one another on the topic of CP.

To conduct these dialogs, all Group 1 members were situated in one classroom and all Group 2 members in an adjacent classroom. Dialogs were conducted via *FirstClass*© IM software, installed on each participant's individual laptop computer. Partners were instructed to conduct a dialog via IM. Their goal, they were instructed, was to try to convince their partners that their position was the better one. These dialogs lasted from 10 to 20 min; they ended when participants indicated they had finished or when no

more time was available. The software saved the dialog for analysis.

Intervention: Collaborative Dialogic Argument Plus Reflection

The argumentation activity took place in three identical phases, each devoted to a different topic. The first of the three phases, described here in detail for illustration, comprised thirteen 40-min classroom sessions occurring twice per week. The activity was introduced to students as a class in learning to become good arguers. The class project was to focus on ColumbiaTown, a new town being formed in the middle United States. The governing board of the new town had to establish rules and laws for the new town, and in particular, the participants' task would be to focus on the new school that was being formed to serve the town's children. The first topic was to be whether attendance at school should be mandatory for all children in the town. Specifically, the following dilemma was introduced:

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

Phase 1 individual assessment. The positions of all participants on this topic were elicited via an individual written questionnaire. The questionnaire asked the participant to indicate their position among the three options, ''Nick should go to the school,'' ''Nick can be taught at home,'' and ''undecided.'' Participants were then asked to indicate how certain they were of their positions on a 6-point scale ranging from *certain* to *not sure at all*. The questionnaire then proceeded to pose the same sequence of probing questions described earlier for the CP-topic individual assessment.

Phase 1 dialogic engagement. Based on responses in the individual assessment, two groups of 14 students each were formed, one group who chose the homeschool option and another group who either chose mandatory townschool ($n = 9$) or were undecided but noted reasons supporting the townschool side ($n = 5$). Within each of the groups, homeschool and townschool, same-gender pairs were formed. To the extent possible, an additional criterion used in forming these pairs was to pair participants who provided different reasons for the same position.

The same-side pairs thus formed remained together until the ''showdown'' segment (see below) of Phase 1. Beginning at the next class session, pairs in one group assembled in one classroom and pairs in the other group assembled in an adjacent classroom. Each pair was assigned to conduct a dialog via IM with a pair from the other room. Their goal, they were instructed, was to try to convince the other pair that their position was the better one. Each pair was further instructed to collaborate with their partner to decide what they wished to say to the opposing pair and, once they were in agreement, to type their response and send it to the opposing pair.

Those in one room were arbitrarily chosen to begin the dialogs by stating and supporting their position, with the pairs in the other room then given the opportunity to respond. Each of the pairs in one room thus continued their dialogs with one of the pairs in the other room, with dialogs lasting an average of 25 min. At the next session, each pair conducted a dialog with a different opposing pair, until each pair had debated every opposing pair—a total of seven sessions. One or two adult coaches circulated each room during these sessions, answering any questions and reminding pairs to collaborate with one another in deciding what response to make.

Participants received no explicit instruction with respect to argumentation. The only exception, used occasionally if it appeared in danger of being violated, was to ''criticize ideas, not people.'' The only implicit instruction was the invoking of a metaphor expressed occasionally to the whole group that in responding they should strive to ''get the ball over the net and back to the other side.''

Phase 1 reflective activity. After four of the seven dialog sessions had been completed, the first reflective activity was introduced at the next session. At these sessions, each pair was given a printed transcript of the preceding session's dialog. (Because sessions took place twice a week, the interval between the dialogic session and the reflective session was 3–4 days.)

In the reflective session, in addition to the transcript of the preceding session, pairs were given one of two scaffold sheets. The ''Other's Argument'' scaffold sheet prompted pairs to examine the opposing

pair's argument and their own counterargument. The "Own Argument" scaffold sheet prompted pairs to examine the opponent's counterargument and their own rebuttal to this counterargument. In each case, they were asked to contemplate what a better counterargument or rebuttal might have been. The adult coaches circulated during these sessions to provide clarifications if needed.

Reflective sessions took place a total of three times during Phase 1, interspersed among the dialog sessions. Specifically, the first Other's Argument session occurred after the fourth dialogic session, the first Own Argument session occurred after the fifth dialogic session, and a combined reflective session in which both scaffold sheets were available occurred after the sixth dialogic session.

Phase 1 preparation and showdown. Following the seven dialogic and three reflective sessions, participants were told at the next session that it was time to begin preparation for a final showdown debate with the opposing side. They were given one session to prepare for the showdown and the following session to conduct it.

To prepare, pairs within their opposing-sides classrooms were regrouped into an A and a B team. One team was encouraged to become "own argument" specialists by reviewing the possible counterarguments the opposition would make to their arguments and the rebuttals they would have ready. The other team was encouraged to become "other argument" specialists by reviewing the possible arguments the opposition would make and the counterarguments they would have ready. Transcripts of the dialogs, the previously completed scaffold sheets, and blank index cards were made available.

For the showdown, the two opposing groups remained in their respective classrooms and communicated with one another via the same IM software but this time with a single computer terminal and Smartboard in each room. To conduct the showdown, students were again reorganized, this time into Team A and Team B (each containing similar numbers of "own argument" and "other argument" specialists), with Team A allocated to be the acting team for the first half of the activity (with Team B advising by transmitting suggestion cards to the acting team), and the roles reversed for the second half of the activity. The entire activity lasted 50 min.

At the session following the showdown, an argument map that had been prepared by the researchers was presented to the assembled class. Different colors, as well as labels, were used to identify argumentive moves as effective (green), less effective or ineffective (red), and neutral (blue or black). A point system was applied to this map, making it possible to declare a winning team.

Phases 2 and 3. Phase 2 began following winter vacation and Phase 3 following spring vacation. Phases 2 and 3 proceeded in an identical way to Phase 1 except for new topics and a slight reduction in number of sessions (one less dialogic and one less reflection session) due to the slightly reduced time available in the school calendar for these two phases.

The topic for Phase 2 was whether students who continually disrupt the classroom should be expelled or allowed to remain. The topic for Phase 3 was whether teachers should receive experience-based pay (with more experienced teachers more highly paid) or all teachers should receive the same pay.

Final Assessment

The final assessment occurred following Phase 3. It was on the same (CP) topic and contained the same components as the initial assessment. In the dialogic activity, all participants argued with the same opponents they had engaged in the initial assessment.

Results

We begin analysis with a comparison of participants' performance on initial and final assessments. This analysis allows us to establish the overall progress in skills that each participant made, assessed in a content domain different from the ones involved in the intervention. In addition to its importance in its own right, this information proves useful in examining progress during the intervention itself. Our initial analytic task, however, before microgenetic examination could begin, was to apply a common dialogic coding scheme to the initial and final dialogs, as well as to the intervention dialogs.

Coding of Electronic Discourse

Each of the 161 electronic dialogs (7 per session at 19 intervention sessions across the three intervention topics and 14 each at the initial and final assessments) was analyzed based on an elaborated version of the scheme used in earlier research (see Felton, 2004; Felton & Kuhn, 2001; Kuhn & Udell, 2003; Udell, 2007). The discourse is segmented into idea units and a code assigned to each unit based on its functional relation to the opponent's immediately preceding utterance in the dialog. Coders blind to the identity of participants or time of the dialog took part in the coding. Interrater reliability (calculated on a

randomly chosen 15% of the intervention dialogs and 100% of the initial and final dialogs, all of which were double coded) indices were 89% agreement on individual units within dialogs and a Cohen's kappa of .87. Coding reliability was examined individually for each of the major categories to be examined here and was found to be in a comparable range. The percent agreement for the Counter-C category (see definition below), for example, was 98% with a Cohen's kappa of .89. An additional coding scheme for meta-level discourse was developed to aid in examination of the intervention dialogs and will be described in presenting the intervention data analysis.

Skills at Initial and Final Assessments

At the initial assessment, dialogs contained a mean of 14.94 coded utterances ($SD = 8.26$). At the final assessment, dialogs contained a mean of 15.06 utterances ($SD = 8.12$), a nonsignificant difference. Mean number of words contained in each utterance increased, however, from 6.83 ($SD = 2.57$) at the initial assessment to 11.70 ($SD = 4.44$) at the final assessment—a significant increase, $F(1, 54) = 25.28$ ($p < .001$).

In comparing the kinds of utterances contained in dialogs at the two assessment times, only those categories from the coding scheme are examined that accounted for greater than 5% of utterances, averaged across dialogs, by the time of the final assessment. These categories are Clarify, Case?, Counter-A, and Counter-C. Percent use of each of the other categories was 5% or below at one or both assessment points. (Meta-level categories are addressed later.) In the comparisons across time reported below, an arcsine transformation was performed to normalize proportions.

Counterargument (Counter-A and Counter-C). Counterargument is the central and most essential skill to be mastered if one is to engage in effective argumentive discourse. Its use reflects awareness of the dual goals of argumentive discourse noted earlier (to weaken the opponent's argument, as well as to gain concessions from the opponent that will support one's own argument). We assess the use of two forms of counterargument. The weaker form, Counter-A (counter-alternative), addresses the other's position and undertakes to reduce its force, but it does so by introducing a new criticism of the other's position, not directly related to the argument the other has just made in support of his or her claim. The stronger, more skilled form of counterargument, Counter-C (counter-critique), by contrast, directly addresses the other's argument, criticizing it in a way that seeks to weaken its force. In the case of the CP topic, for example, if the other's argument is "CP will keep dangerous criminals off the street," a common Counter-C is "Life imprisonment can do that as well," whereas a common Counter-A is "Sometimes innocent people are executed."

In analyzing change in counterargument usage from initial to final dialogs, we examined three indices of counterargument skill. The first is the proportion of utterances a participant made that were counterarguments of either type. The second is the proportion of utterances that were coded as Counter-Cs. The third is the proportion of utterances coded as counterarguments that were Counter-Cs. Thus, we look both at mastery of counterargument itself and then at indications of the extent to which counterarguments are being used most effectively.

All three of these indices showed significant increases from initial to final assessments. The percentages were transformed using an arcsine transformation. Overall Counter usage increased from an average of 6.34% ($SD = 7.66$) of utterances at the initial assessment to 26.53% ($SD = 17.27$) at the final assessment, a significant change over time, $F(1, 27) = 31.75$, $p < .001$, $\eta_p^2 = .540$. Counter-C usage increased from 4.50% ($SD = 7.09$) to 21.37% ($SD = 14.61$), $F(1, 27) = 19.89$, $p < .001$, $\eta_p^2 = .424$. Finally, the percentage of Counters that were Counter-Cs increased from 31.25% ($SD = 45.45$) to 63.79% ($SD = 37.75$), $F(1, 27) = 7.27$, $p = .012$, $\eta_p^2 = .212$.

As important as these group trends are patterns of individual change, especially as we wish to relate such patterns to patterns of individual change that participants show during the intervention itself. At the initial assessment, only 36%—10 of 28—participants made any Counter-C arguments. At the final assessment, all but one did so (a significant change, McNemar test, $p < .001$). At the initial assessment, 50% (14 of 28) participants made *any* counterarguments at all, but all did so at the final assessment (also a significant change, $p < .001$).

Also of interest are the proportions of participants who showed increases in their overall Counter and Counter-C arguments from initial to final assessments—a majority did so, 85% in the case of overall Counters and 78% in the case of Counter-Cs. Only 1 participant showed no Counter-C usage at either assessment. Of the remaining participants who failed to show improvement in Counter-C usage, all showed some slight usage at each assessment.

Rebuttal. A further important aspect of counterargument is the extent to which discourse participants are able to maintain focus and consistency in intent and execution that enables them to sustain the critique of one another's arguments. Once the initiator of

a claim responds to an opponent's counterargument by countering this counterargument, we have what is commonly referred to as a rebuttal. In our analyses of participants' dialogs, we looked at both frequency of rebuttals and length of rebuttal strings, that is, over how many contributions to the dialog a sequence of Counters is maintained.

Participants in the present study showed little initial skill in rebuttal. Only 6 of 28 (21%) made any rebuttal utterances at the initial assessment. At the final assessment, this number increased to 19 (68%)—a threefold and significant increase ($p < .001$). Consistent with this increase, mean number of rebuttals across participants increased significantly from 0.39 ($SD = 0.88$) at the initial assessment to 1.82 ($SD = 2.06$) at the final assessment, $F(1, 26) = 10.51$, $p = .003$. Overall frequencies are presented in this case, rather than percent usage, because production of rebuttals is affected by the opportunities available to make them. In other words, a participant cannot make a rebuttal if the opponent has made no counterargument. We therefore also calculated the percentage of rebuttal opportunities that were taken advantage of, that is, the number of rebuttals an individual made relative to the number of opportunities for rebuttal offered by the opponent. This percentage increased from an average of 17.65% ($SD = 35.15$) at the initial assessment to 43.79% ($SD = 36.48$) at the final assessment. This change, note, however, is confined to a reduced sample, the 12 participants who had any opportunities for rebuttal at the initial assessment (all did so at the final assessment), and this sample size is not large enough for statistical test. Interpretation of group data on increasing rebuttal skill thus remains limited until counterargument skill reaches a level at which all participants have available numerous opportunities to make rebuttals.

We looked finally at the length of rebuttal strings, whenever rebuttals were made. A length of 1 represents a sequence consisting of assertion – counterargument – counterargument; a length of 2 represents a sequence consisting of assertion – counterargument – counterargument – counterargument. Mean length rose only slightly from 1.83 at the initial assessment to 2.13 at the final assessment (a statistically nonsignificant increase). These numbers show a typical successful effort on the part of a maker of an initial counterargument to not allow to stand as successful the opponent's move to weaken this counterargument (length of 2, i.e., assertion – counter – counter – counter).

Exposition (Clarify and Case?). Clarify, the most common utterance category, includes all those utterances in which the participant asserts or elaborates his or her own arguments. Although in previous work (Kuhn & Udell, 2003) usage in this category has declined as counterargument usage increased, participants in the present study overall showed a constant percent usage of Clarify—an average of 21.6% ($SD = 10.1\%$) at the initial assessment and 21.7% ($SD = 14.2\%$) at the final assessment. Because overall length of dialogs (reflected in number of utterances) remained constant, as reported earlier, this finding indicates absence of change in either number or proportion of utterances in the Clarify category.

Case? is another category that in past research has shown different patterns of change in different samples, with the suggestion that its use may increase temporarily but then diminish as other argument strategies become available (Felton, 2004; Felton & Kuhn, 2001). In Case?, the participant poses a situation to the other and asks the other to respond. In discussing the CP topic, for example, one member of the dyad might ask the other how she would feel about CP if the murder victim were a family member. The use of Case? is effective when it poses the situation with the intent of securing concessions from the other and in this respect may be seen as a stepping stone or alternative to direct counterargument. In the present analysis, the average percent use of Case? increased from 1.54% ($SD = 3.10$) at the initial assessment to 6.84% ($SD = 11.68$) at the final assessment, $F(1, 27) = 13.18$, $p = .001$, $\eta_p^2 = .328$.

Individual Argument

Although our purpose in the present work is to examine the development of dialogic argument skills in their own right, we did obtain evidence that increasing dialogic skill leads to improvements in individual (nondialogic) argument. Such evidence corroborates our own and others' earlier studies of individual argument skills (Kuhn, Shaw, & Felton, 1997; Reznitskaya et al., 2001) in which such positive effects of practice in dialogic argument on individual argument have been observed (allegedly due to enhanced awareness of the ''missing interlocutor'' implicit in nondialogic arguments).

The coding scheme developed by Kuhn et al. (1997) was used for assessment of individual CP arguments. In particular, we examined the quality, quantity, and sidedness (two- vs. one-sided) of initial and final individual arguments for or against CP. As 1 participant was away for medical reasons at the final individual assessment, these analyses are based on a sample size of 27. Level 1 reasons in this scheme offer genuine justifications for or against the use of CP, whereas Level 2 reasons offer only functional arguments regarding the criteria for use or nonuse of CP

(e.g., ''only if it's a serious crime''), without addressing justification of the practice itself, and Level 3 reasons are entirely nonjustificatory (e.g., appeal to precedent or authority). (See Kuhn et al., 1997, for the full scheme.)

At the initial assessment, a mean of 0.56 ($SD = 0.93$) different Level 1 reasons were produced compared to a mean of 1.33 ($SD = 1.00$) at the final assessment—a significant increase, $t(26) = 2.95$, $p = .007$. (Repetitions of the same argument code did not increase an individual's score.) Of the 27 participants, 25 added at least one new Level 1 reason at the final assessment. Total number of different reasons (of all levels) increased from a mean of 3.07 ($SD = 1.59$) to a mean of 4.22 ($SD = 1.80$)—a significant increase, $t(26) = 2.37$, $p = .026$. The number of participants producing two-sided arguments (citing reasons both for and against CP), either initially or when elicited (''Can you imagine someone having another position that is different from your own? What would it be?'') increased from 17 to 22 across the two assessments.

Dialogic Skill Development During the Intervention

Analysis of the discourse patterns during the three phases of the intervention is based on a sample size of 28. We divide the presentation of results into two sections, the first section reporting on trends in discourse performance itself and the second section examining trends in the use of meta-level discourse that seeks to evaluate, question, monitor, direct, or otherwise manage this performance.

Overall patterns of change at the group level. For these analyses, the dialog sessions for each topic were collapsed into beginning, middle, and final segments (with two sessions combined in each category except for the middle category for Topic 1, which included three sessions). A participant received the same score as his or her (same-side) partner for a particular session (and the pair taken as the unit of analysis where appropriate in statistical analyses). Dialogic sessions remained fairly stable over time with respect to length. Mean number of utterances per session were 42.12 ($SD = 8.22$) for Topic 1, 37.36 ($SD = 9.21$) for Topic 2, and 42.71 ($SD = 10.80$) for Topic 3. Mean length of utterances increased only slightly from 8.11 ($SD = 6.17$) for Topic 1 to 9.55 ($SD = 8.10$) for Topic 2 and 9.17 (8.52) for Topic 3. Mean word length is overall higher, however, for the Counter-C category, a major focus of our interest, and increased across topics, from 11.29 ($SD = 6.85$) for Topic 1 to 13.75 ($SD = 8.61$) for Topic 2 to 15.74 ($SD = 9.45$) for Topic 3.

Counterargument. In Table 1, there appear the mean percentages of all utterances that were coded as Counter (top row) or Counter-C (second row). The final row shows the mean percentages of Counters that were coded as Counter-C. As reflected in the leftmost column (beginning of Topic 1), the mean percentage of utterances that a participant devotes to counterargument increases substantially, to 24.43%, from the earlier reported mean level of 6.34% at the initial (CP) assessment, where participants argued alone against their opponent rather than with a partner as they do here. Similarly, the mean percentage of Counter-Cs increased from the initial CP assessment, from the earlier reported 4.50% when working alone to the 13.87% shown here. These changes are suggestive of the advantages provided by collaboration with a partner as well as what we have observed in past work to be the motivating and focusing effects of working toward the social objective of the showdown.

As seen in Table 1, however, performance continues to improve during Topic 1 but does not increase further in these respects after the first topic. Based on pairs as the unit of analysis, for Topic 1, gains (from beginning to end) are significant for overall Counters, $Z = -2.605$ (Wilcoxon test), $p = .009$, effect size $r = -.696$, and for Counter-Cs, $Z = -2.166$, $p = .03$, effect size $r = -.579$. For Topic 2, gains are significant only for Counter-Cs, $Z = -2.355$, $p = .019$, effect size $r = -.629$ (for overall Counters, $Z = -1.789$, $p = .074$), and for Topic 3, no trends over time are significant (Counter-C: $Z = -0.722$, $p = .470$; overall Counters: $Z = -0.659$, $p = .510$). Note finally that when we

Table 1
Mean Percent Usage of Counterargument and Counterargument-Critique (Counter-C) Across the Three Topics

	Topic 1		Topic 2		Topic 3	
	Beginning	End	Beginning	End	Beginning	End
Total Counters	24.43 (8.50)	32.30 (10.49)	16.19 (5.80)	20.77 (8.55)	20.15 (8.98)	18.70 (8.17)
Counter-C	13.87 (4.89)	20.62 (9.49)	9.28 (4.99)	14.31 (8.17)	10.37 (5.74)	9.32 (7.75)
Percentage of Counters that are Counter-C	58.22 (15.26)	63.30 (17.07)	57.29 (29.52)	61.16 (25.28)	50.64 (27.35)	48.23 (29.02)

Note. Standard deviations appear in parentheses.

compare the percentages in Table 1 to those for the final CP assessment reported earlier (26.53% for Counters and 21.37% for Counter-Cs), it is apparent that gains made during the intervention are maintained when participants return to dialogic arguments against an opponent without the aid of a collaborative partner.

Rebuttal. We do not undertake statistical analysis of changes in percentage of rebuttal opportunities successfully met due to the initial scarcity and only gradual increase in the availability of such opportunities over time, as noted earlier in the report of rebuttal usage at initial and final (CP) assessments. By the time of the third intervention topic, however, rebuttal skill had been attained to the extent that all pairs showed some rebuttal usage during this topic and 41.52% of rebuttal opportunities were successfully taken advantage of (with all participants having such opportunities)—a percentage comparable to the 43.79% noted earlier observed at the final (CP) assessment. The mean length of Topic 3 rebuttal chains was 2.03 (range = 1–7) comparable to the 2.13 (range = 1–5) observed at the final (CP) assessment. An example of a successful rebuttal chain (of Length 2) is presented in Table 2 for Topic 3 (teacher pay).

Individual change patterns. Although the preceding results are suggestive, they do not tell us to what extent individual patterns of performance conform to group trends portrayed in Table 1. For this purpose, we looked at individual change patterns. We focus on the Counter-C indicator, as our most important index of growing skill in argumentive discourse. For purposes of examination, we also divide participants into two groups, the 21 who had shown progress from initial to final CP assessment (based on Counter-C usage) and the 7 who had not.

In Figures 1a and 1b, we present two examples of individual change patterns among the majority group of 21 who showed increase in Counter-C usage from initial to final assessment. In one case (Participant 25), the individual shows no Counter-C ability at the initial assessment but then, in collaboration with successive partners, exhibits some Counter-C usage in engaging with the three topics; at the final (CP) assessment, again working alone, she is now able to produce some Counter-C without a partner's collaboration.

Another participant (Participant 10, Figure 1b) does show some Counter-C usage when working alone at the initial assessment, continues to display it to a variable degree during the intervention, and then increases this usage near the end of the intervention and in transferring back to the solo context. In both cases, then, the social collaboration experience appears to have been fruitful, and the level of performance does not revert to its original level when the social context is removed. These cases illustrate patterns most common among the groups of 21 who progressed from initial to final CP assessment.

Among the 7 participants who did not show any indication of progress from initial to final CP assessment, 4 showed equivalent performance from initial to final assessment and 3 declined slightly, with performance remaining at a low level in all cases (and 0 in one case, shown here). In all these cases, more such usage appeared during the intervention while the participant was working collaboratively, but the participant did not transfer it to the solo context of the final CP assessment. One of these cases appears in Figure 1c (Participant 28).

Within the three topics, patterns of individual performance resemble those of the three illustrative participants in Figures 1a to 1c. Variability is the norm, rather than consistent upward change, within each topic, and performance achieves its maximum levels during the first topic (as also reflected in Table 1)—an unanticipated finding we return to in the discussion of our findings. Despite this variability, there appears within the intervention sessions, at least for the 21 ''progressing'' participants (the group defined by progressive change across CP assessments), a common trend of progress within a topic from the beginning to final sessions of engagement with that topic. To quantify this trend, we defined a pattern as upwardly progressing over the beginning–middle–end phases of engagement with a topic if performance during the second was higher than the first and performance on the third higher than the second. We defined a pattern as mixed upward if performance at the final phase was higher than at the beginning phase, but the two intermediate links (to the middle point) reflected either plateau or decline.

Using these definitions, among the 21 progressing participants (the group defined by progressive

Table 2
Example of Rebuttal Chain for Topic 3 (Teacher Pay)

Pair A: WE THINK THAT THEY SHOULD GET THE SAME PAY, BECAUSE SOME TEACHERS MAY LEAVE (*coded Clarify*)
Pair B: but the teachers won't leave if they think they get more pay next year + It will make them want to stick to the job *(coded Counter-c)*
Pair A: THE TEACHER'S WITH LESS EXPERIENCE WON'T THINK IT'S FAIR. BECAUSE NEXT YEAR IS A WHILE AWAY (*coded Rebuttal*)
Pair B: then they can attract experienced teachers (*coded Rebuttal*)

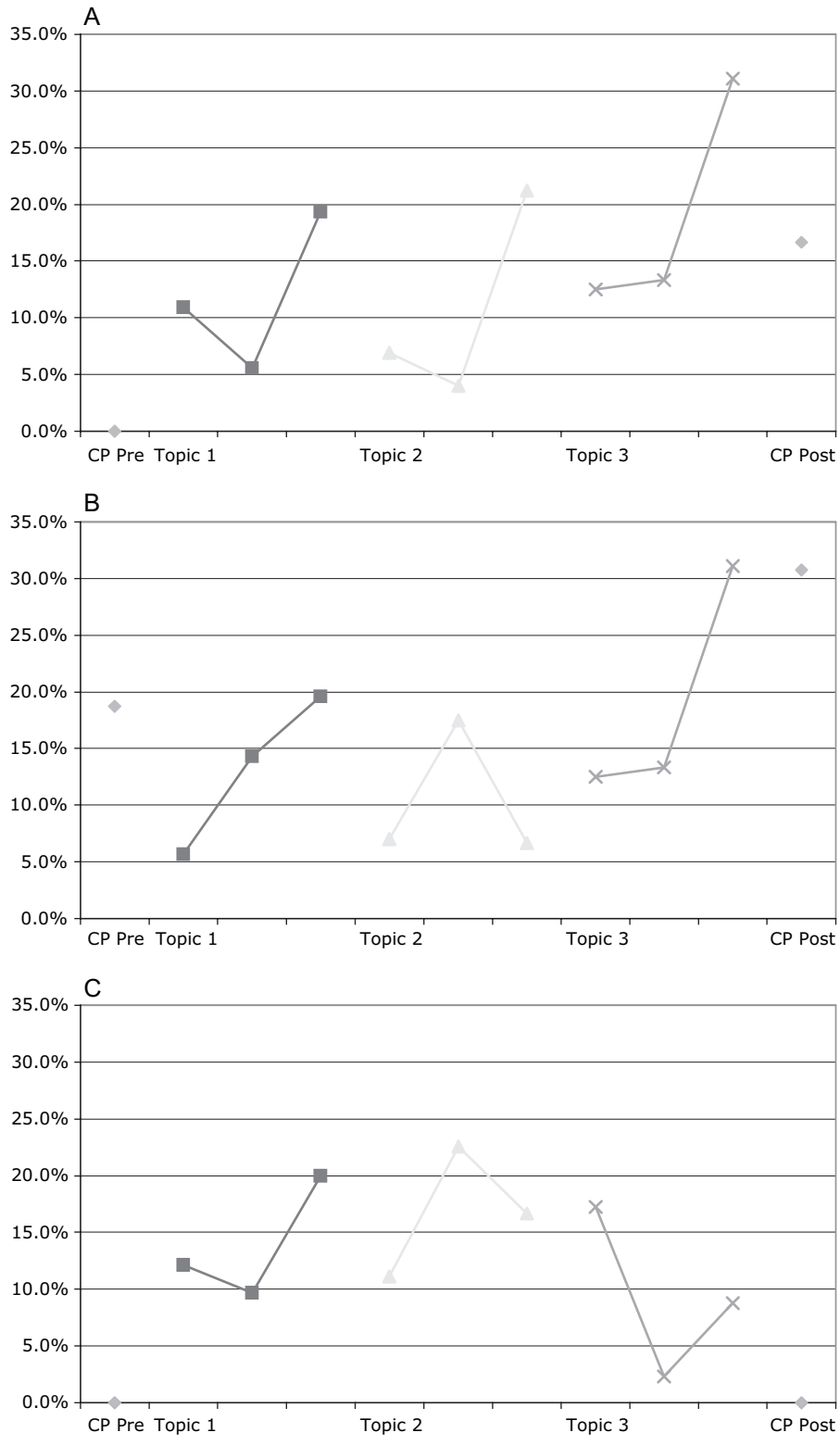


Figure 1. Percent Counter-Critique usage for discourse use across topics: (a) Participant 25, (b) Participant 10, and (c) Participant 28. CP = capital punishment.

change across the two CP assessments), the majority of patterns were in one of these two categories (upward or mixed upward)—specifically 73%. In contrast, among the remaining 7 participants who did not show progress across the two CP assessments, this percentage was 48% (roughly what would be expected if change were in no consistent direction). Although sample size is too small for meaningful statistical analysis, these findings are of interest in suggesting a link between patterns of performance during the intervention and at the initial and final (CP) assessments.

Meta-Level Development During the Intervention

Consistent with the questions we initially identified as central to the present work, our objective in this phase of the analysis was to examine the extent to which participants showed progress in recognizing the goals of argument, as distinguished from progress in implementing successful argumentive strategies. We expected progress of the former type to be most visible when participants collaborated with a partner toward a collective goal of defeating the opposing side in the showdown activity (devised to highlight this goal). We did not have resources in the present study to record the conversation between collaborating partners during the dialogs as they formulated the responses they would make to their opponents. These responses themselves, however, became a matter of record and could be examined with respect to what evidence they showed of meta-level understanding of argumentive goals and strategies. Meta-level statements we defined as statements *about* the dialog in contrast to statements that constitute the dialog. Omitted from this analysis are utterances that were meta-level but procedural in nature, for example, requesting or receiving clarification regarding task instructions or software issues and did not pertain to the dialog itself.

To further examine the nature of these meta-level statements, we developed the coding scheme shown in Table 3. The subcategories of meta-level statements appearing in Table 3 are listed in descending order with respect to their centrality to understanding of the goals and purposes of argument together with their

Table 3
Types of Meta-Level Statements

Type	Definition	Overall percent usage	Examples
Meta-directive	About the dialog, with a directive to opponent	7.59	—please state your argument —give us some reasons —give us proof and reason —Then prove it . . . —you need to give facts instead of opinion —STAte why —LETS GET BACK ON TOPIC —Please answer my question —please answer 2 our arguments!
Meta-argumentation	About the dialog, without an explicit directive to opponent	2.99	—you have not answered —u havent given good answers + so we r gonna keep askin it —that's opinion not a fact —how do you know —We're not talking about that
Meta-argument	About respective positions rather than the argumentation process	4.85	—thats our point —thats what WERE arguing for! —that is our main argument —Yeah thats what weve been trying to say —its the same argument except with some modifications —look we agree on something! —you guys agreed with us before —we never said that
Meta-comprehension	About comprehension	0.82	—You are not making sense —we don't understand —I don't get what you say man

Table 4
Mean Percent Usage for Total Meta and Meta-Directive Across Topics

	Topic 1			Topic 2			Topic 3		
	Beginning	Middle	End	Beginning	Middle	End	Beginning	Middle	End
Progressing participants									
Meta-directive	4.62 (3.90)	5.84 (5.36)	4.91 (3.64)	9.51 (6.49)	14.84 (9.59)	8.43 (7.69)	11.20 (6.28)	10.70 (9.16)	7.55 (4.77)
Total meta	15.27 (5.53)	13.39 (8.92)	11.48 (5.67)	16.64 (7.56)	25.00 (12.48)	16.84 (10.18)	25.18 (12.29)	24.21 (12.37)	21.57 (7.24)
Nonprogressing participants									
Meta-directive	6.03 (3.20)	4.82 (2.45)	6.12 (4.99)	10.15 (5.30)	6.56 (7.72)	7.43 (6.09)	10.27 (7.23)	18.09 (13.39)	9.06 (4.62)
Total meta	13.50 (5.09)	13.05 (9.37)	12.12 (9.38)	15.77 (7.17)	12.17 (5.31)	14.93 (8.53)	19.31 (8.38)	24.85 (15.72)	22.00 (10.56)

Note. Standard deviations appear in parentheses.

overall frequencies of usage. After using a portion of the data to develop the scheme, two of the authors independently coded one third of the data, achieving a percent agreement of 82% for identification and categorization of meta-level statements, with a Cohen's kappa of .69. Differences were resolved by discussion, and all remaining statements were coded by one of these two coders.

Overall change at the group level. To conduct an analysis of change over time in the various kinds of meta-level statements, as well as overall usage of meta-level statements, we again examined separately the participants in two groups, the majority group ($n = 21$) who had shown positive change in the initial to final CP assessment, and the minority group ($n = 7$) who showed no indication of such change (''progressing'' and ''nonprogressing'' group, respectively). In Table 4, there appears for each group mean percentages of usage across time of utterances of a meta-level type. (The time intervals shown are the same as those in the earlier reported analyses in Table 1.) As seen in Table 4, for the majority (progressing) group, the mean percentage of meta-level statements increased from 15.27% at the beginning of Topic 1 to 21.57% at the end of Topic 3, a significant increase, with gains for the minority group comparable. These gains are significant for the sample as a whole, $Z = -3.552$ (Wilcoxon test), $p < .001$, effect size $r = -.671$, as well as for the progressing group examined individually. (Note, however, that in contrast to the analysis of Table 1 data, the individual, rather than the pair, must be the unit of analysis, as pair composition changes across topics. Given the lack of independence between members of a pair, significance levels should therefore be interpreted conservatively.)

Change over time was also examined for each of the subcategories in Table 3. Change is significant only for the first of the subcategories—Meta-directive. For the sample as a whole, $Z = -2.619$ (Wilcoxon test), $p = .009$, and effect size $r = -.495$ (although the same caution noted earlier applies).

These gains, however, were not maintained when participants returned to arguing against an opponent without a partner in the final (CP) assessment. Meta-level coding of the CP dialogs showed a steady rate of meta-level usage at the two points, for example, for meta-directives, 5.6% at initial and 5.8% at final CP assessment.

Individual change patterns. Examination of individual patterns of performance with respect to meta-level usage confirms that the pattern of increasing meta-level usage, in particular the Meta-directive category, is consistent across participants. Only 2 of

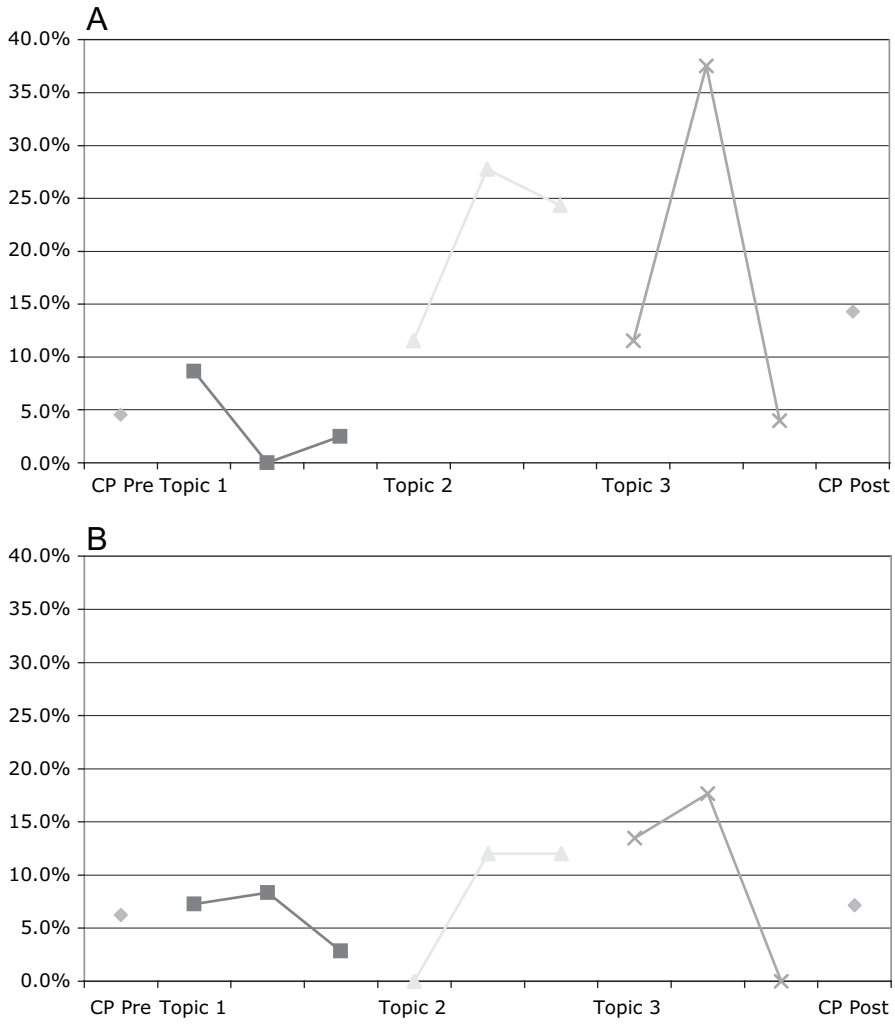


Figure 2. Percent meta-directive usage for meta-level discourse across topics: (a) Participant 19 and (b) Participant 11.

28 participants in fact fail to show such a pattern. Two typical examples appear in Figures 2a and 2b (Participant 19) and (Participant 11). As seen there, the progress reflected in Figure 2b is less dramatic than that in Figure 2a. In both cases, however, there appears the typical pattern of a decline in Meta-directive usage when the participant no longer works with a partner in the final (CP) assessment.

Implications of meta-level change. In summary, almost all participants showed increasing meta-level usage during the course of their collaborative argumentation activities. At the final dialogic assessment, no longer working collaboratively, three quarters of participants (21 of 28) showed enhanced argumentation skill (indexed by percent use of Counter-C). Did this increase in meta-level discourse about argumentation reflect understanding that contributed to the enhanced skill observed in the conduct of argumentation? Our theoretical assumptions regarding the importance of meta-level regulation of behavior suggest a positive answer, but it is not straightforward to identify strict causal evidence of such an effect. Because meta-level discourse emerged and increased over time in almost all participants, we cannot compare those who did and did not engage in meta-level discourse. Furthermore, participants worked in pairs during the intervention, and we cannot be sure that the meta-level discourse directed to the opponents was contributed to equally (or at all) by each member of the pair (and would therefore be expected to affect the individual skill assessment of both members of the pair).

In search of some further evidence of the contributory role of meta-level processes, we examined pairs who showed more versus less than the typical proportion of meta-level discourse to assess whether these two groups would differ in argumentation skill level. For this analysis, we chose Topic 2, when meta-level

discourse was most likely to have emerged, after participants had become familiar enough with the activity to begin to reflect on how best to carry it out. We divided the Topic 2 pairs (partners differed, recall, for the other two topics) into those who showed relatively high proportions of statements in the meta-directive category (15% or more) and those who showed lower proportion of statements in the meta-directive category (less than 15%). Of the 14 pairs, 8 fell into the high category and 6 into the low category. For each pair, we calculated the mean of the 2 pair members' proportion usage of Counter-C at the final (CP) dialogic assessment. Among the 8 high meta-directive pairs, 7 showed Counter-C usage equal to or greater than 15%. Among the 6 low meta-directive pairs, only 1 pair showed Counter-C usage equal to or greater than 15%—a significant difference (Fisher exact test, $p = .026$). Although this comparison involves small numbers and should only be regarded as suggestive, it does offer some evidence of a link between meta-level discourse and argumentation skill.

Showdown

Although we did not formally analyze performance in the showdown preparation sessions, we did undertake the same discourse analysis of the three showdown transcripts as was performed on the intervention dialogs. Here, we saw that the focused group effort produced a higher level of skill than that observed during the dialogs between pairs. The overall percentages of showdown Counter-C statements were 58% for Topic 1, 61% for Topic 2, and 70% for Topic 3. To illustrate the level of performance that participants achieved as a group, we present in the Appendix the (slightly edited) showdown debate produced for the homeschool topic in the form of the argument map presented to students at the session following the showdown for this topic. The Appendix contains a black and white version (with effective moves in bold and ineffective in dotted lines) rather than the colored version (with green and red boxes) that students saw.

Discussion

Our findings highlight the value of the microgenetic method in providing a clearer and fuller picture of the changes that occurred among our participants than would have been evident from the pretest–posttest assessments alone. We therefore discuss our results in the reverse order in which they were presented, beginning with the microgenetic data on meta-level change and then proceeding to change at the strategic level and finally to the pre- and posttest assessments of first dialogic and then individual argument skills.

On virtually all the dimensions we examined, some participants were identified who failed to show progress. This group was smallest in number, however, on the dimension of meta-level discourse. Of the 28 participants, 26 showed a pattern that involved increasing usage of meta-level discourse, in particular meta-directives to the opposing pair. As elaborated earlier, we see this meta-level discourse as significant in indicating at least implicit understanding of the functions and objectives of argument. When participants say to their interlocutors, "Give us some reasons" or "You have not answered," they are displaying implicit understanding of what contributors to a dialog need to do in order to engage effectively in argumentation.

It is notable that these increases on the part of almost all participants were evident only when participants were engaging in the dialogic task with a collaborating partner. When participants again worked without a partner (at the final CP dialogic assessment), these increases were no longer evident (as reflected in comparisons of initial and final CP dialogs). In the collaborative context, participants had the cognitive and social support of one another in formulating their contributions to the dialog and indeed needed to collaborate in order to perform their task. They also shared the social goal of prevailing over the opposing side in the showdown. Neither of these conditions existed when they argued individually with an opponent on a different topic, one that they had discussed only once before and without a particular objective. We cannot say with certainty whether under the latter conditions participants lost sight of argumentation objectives that they were aware of in the collaborative condition. We see as a more likely interpretation, however, that they simply perceived no need to express them in meta-level talk in the solitary condition. They had a task to do (to debate the CP topic with an opponent), and there was no one they needed to consult with as to how best to carry out the task, so they simply got on and applied their skills to the task, with little comment on their own or their opponent's performance. If so, our findings point to the importance of social context and goal structure in influencing performance—a conclusion also reached by Nussbaum (2005) in a study of argumentation in a college-student population. The association we observed between extent of meta-directive discourse and later individual argument skill nonetheless supports the view that this meta-level discourse was productive with respect to skill development.

Turning to the strategic skills themselves, it is notable that, in contrast to what was the case for the meta-level discourse just discussed, positive change over time did not diminish when participants return to the nonsocial context of working without a collaborating partner and without a social goal. Counterargument and rebuttal, which lie at the heart of argumentation, increased in frequency among three quarters of the participants from initial to final CP assessment, and all participants showed some skilled counterargument (Counter-C) at the final assessment (compared to only a third at the initial assessment). The increase in frequency of skilled counterargument (Counter-C), to 21.37% of discourse, is consistent with our earlier studies involving similar methods (Felton, 2004; Kuhn & Udell, 2003; Udell, 2007), and the amenability of argumentation skill to increased usage with exercise and scaffolding of various kinds is also consistent with other lines of research involving both children (Anderson et al., 2001; Nussbaum, 2003; Reznitskaya et al., 2001) and young adults (Clark & Sampson, 2005; Glassner & Schwarz, 2005; Nussbaum, 2005).

An unexpected finding, revealed by microgenetic analysis, is that strategic-level (although not meta-level) gains were concentrated in the first third of the intervention activity (Table 1). We cannot, of course, be certain that the gains documented at the final (CP) assessment would have been the same had the participants not had the further experience of engaging in the second and third phases of the intervention. Still, we are left with the question of why they did not show further strategic gain during the second and third phases.

Although numerous explanations are plausible (e.g., motivational factors or ceiling effects), the explanation we favor derives from our going back to the dialog transcripts and comparing their content across the three intervention topics. Although we did not intend it, there may have been a subtle difference across topics. The first topic, the homeschool topic, was characterized by well-defined opposing positions (the protagonist in the dilemma is either homeschooled by his parents or attends the townschool), either of which could plausibly lead to a satisfactory outcome. The second topic, school expulsion, also had well-defined opposing positions (the misbehaving students are removed from the school or are not). In this case, however, neither alternative can be expected to lead to a satisfactory outcome for the misbehaving students. In the words of a number of our participants on both sides of the issue, ''That won't solve the problem.'' Participants on one side noted that these students would continue to disrupt if not removed, whereas participants on the other side often noted that the students would just cause the same disruption if removed and sent elsewhere to school. The latter group thus devoted considerable dialog to proposing other solutions (e.g., bringing in a social worker or counselor to work with the misbehaving students) that would resolve the problem.

What we may be seeing here, we believe, however, is a shift in goal structure, driven by the topic content, from the classic *argument* goal structure (having one's position prevail over that of an opponent) to the *dilemma resolution* goal structure of resolving the underlying problem that the two opposing sides address. Our dialogic coding scheme, which classifies the function, not the content, of utterances does not provide direct evidence of such a shift across topics, and it must remain speculative. Our discourse analysis of the three showdowns, however, showed that even if participants were diverted from counterargument/rebuttal in their paired dialogs during the later topics, their skill in counterargument was not compromised when they proceeded to the focused context of the showdowns.

The appearance of a shift in focus during the paired dialogs becomes even more pronounced in examining dialogs on the third topic, teacher pay. A subgroup of participants on the experience-based-pay side of the issue during the course of the dialogs identified as their position a third alternative, merit-based-pay, acknowledging weaknesses of the original experience-based-pay position. Some of their dialog was thus diverted to identifying and elaborating this new position and away from counterargument against the opponents' position.

Should we regard such shifts as advances, as refusals to accept the opposing positions as defined and the seeking of a ''third way?'' Such an interpretation is consistent with the ideas of philosophers such as Gilbert (1997), who defines a model of what he calls ''coalescent argumentation,'' the goal of which is to enrich group understanding rather than pit sides against one another—the sort of model that has been proposed by some psychologists as a possibly more fruitful model for instructional interventions (Anderson et al., 2001; Nussbaum, 2005).

While appreciating Gilbert's model, as developmentalists, we are inclined to see the value of older children and adolescents gaining mastery of traditional argumentation structure, as a foundation for acquiring skill in more complex argument forms. In previous research (Kuhn & Udell, 2007), we have observed young adolescents rely on what we called an avoidant pattern in addressing very simple arguments, stating, for example, ''I'd just play both,'' in response to a question about the merits of choosing

soccer versus basketball as one's sport. Mastery of the traditional argument – counterargument – rebuttal structure involved in debating two opposing positions, in our view, is a component of, and hence a necessary foundation for, engaging in more sophisticated forms of "coalescent" argumentation. In its absence, inexperienced arguers may find themselves uncertain as to whether a potential contribution to discourse that they may be contemplating is simply an unanchored addition to the current talk about the topic ("Here's another thing that might be relevant") or has an identifiable function to perform in the discourse.

We note, finally, that advances in dialogic argumentation skill did show evidence of influence on the individual written arguments collected as part of the final assessment. Although demonstrating such effects was not the focus of the present work, it is of course a result one would hope for, as well as one that has been reported in previous studies of the effects of discourse on individual argument skills and/or conceptual understanding in both children younger than those in the present study (Anderson et al., 2001) and in adults (Nussbaum, 2005), as well as in our own earlier work with the age group studied here (Kuhn et al., 1997). Note that new, more adequate reasons in support of his or her position on CP are not ones a participant would have had an opportunity to be exposed to from an external source. The most likely interpretation of their appearance, in our view, is that it is the cognitive effort of considering multiple views and generating counterarguments and rebuttals with respect to them, as well as the benefit noted earlier of increased awareness of the "missing interlocutor," that is responsible for the shifts observed toward more adequate, higher level reasons to justify one's position and away from less adequate, lower level ones.

Because our major interest in the present work was on the development of dialogic argument skill, we included as the main focus of our initial and final assessments the dialogic argument skill that a participant displayed in the absence of the social support that characterized the intervention. As a result, participants had some, albeit very limited, dialogic experience with the transfer (CP) topic, and this experience could have contributed to the improvements observed in their ability to produce an individual argument in support of their position on the topic. Evidence against such an effect, however, comes from a study by Udell (2007), with the CP topic among a slightly older adolescent group, in which no improvements occurred in either dialogic or individual skills as the result of simply an initial assessment of these same skills several months earlier. In future work, a fuller assessment of individual argument skills, and how they are affected by advances in dialogic and meta-level argumentation skills, is warranted to further elucidate the suggestive gains observed in this study.

In conclusion, the microgenetic method employed here has shown that argumentation skills are amenable to development with engagement and practice in a scaffolded context yet without direct instruction. In light of the earlier noted weaknesses in argument skills reported by educators among students at the high school and even college level, this finding alone is notable, especially given that our results also support the idea of dialogic argument as a developmental pathway to individual argument skills of the sort emphasized in academic contexts.

In addition, however, the present work we believe points to the importance of meta-level awareness and understanding of argumentation rather than simply the implementation of argument strategies at the performance level. Understanding and appreciating the value of argument is an epistemological achievement (Kuhn & Park, 2005). In the broad sense, this means seeing the point of argument and coming to appreciate it as the ultimately most powerful means of influencing others' thinking. Only in so doing will one be disposed to invest significant effort in developing the skills that argument entails. Children need to learn that argument is more than something to be avoided. This understanding is not intuitively given. Certain cultural values (to "live and let live" and to show tolerance toward others' views and avoid risk of offending) work against it, and only a more advanced level of epistemological understanding supports it (Hofer & Pintrich, 1997, 2002; Kuhn, 2005; Moshman, 2005).

In the narrower sense, this meta-level understanding entails understanding of the objectives of argument and specifically of the relevance of the other person's position. If the opponent's position is not relevant, the process through which one achieves victory over the opponent cannot be regarded as one of argument. Kuhn and Udell's (2007) study of children's and adults' arguments found, in both groups (although to a greater extent among children) and in multiple (open- and closed-ended) formats, a preference for articulating one's own position over examining and addressing the other's position. As noted earlier, this finding supports the view that the developmental challenge may lie more in understanding the relevance of argument strategy than in executing it. If so, the present work suggests that such understanding is amenable to development.

Leitao (2003) showed college students a series of short essays together with additional sentences of counterargument and rebuttal to the main argument, which could potentially be inserted into the essay.

Students were divided in their opinions as to whether adding such counterarguments and rebuttals would improve the persuasiveness of an essay. Such findings highlight the importance of meta-level development with respect to argumentation. Although tracking this meta-level development is less straightforward than tracking development at the strategic level of performance, both seem essential.

References

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

Andriessen, J. (2006). Collaboration in computer conferencing. In A. O'Donnell, C. Hmelo-Silver, & G. Erkens (Eds.), *Collaborative learning, reasoning, and technology.* (pp. 197–231). Mahwah, NJ: Erlbaum.

Andriessen, J., Baker, M., & Suthers, D. (Eds.). (2004). *Arguing to learn: Confronting cognitions in computer-supported collaborative learning*. New York: Springer.

Bell, P., & Linn, M. (2000). Scientific arguments as learning artifacts: Designing for learning from the Web with KIE. *International Journal of Science Education*, *22*, 797–817.

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

Chinn, C. A. (2006). Learning to argue. In A. O'Donnell, C. Hmelo-Silver, & G. Erkens (Eds.), *Collaborative learning, reasoning, and technology.* (pp. 355–383). Mahwah, NJ: Erlbaum.

Clark, V., & Sampson, D. (2005, June). *Analyzing the quality of argumentation supported by personally-seeded discussions*. Paper presented at the Computer-supported collaborative learning conference, Taipei, Taiwan.

Damon, W. (1984). Peer education: The untapped potential. *Journal of Applied Developmental Psychology*, *5*, 331–343.

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

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

Gilbert, M. (1997). *Coalescent argumentation*. Mahwah, NJ: Erlbaum.

Glassner, A., & Schwarz, B. (2005, June). *The role of floor control and of ontology in argumentative activities with discussion-based tools*. Paper presented at the Computer-supported collaborative learning conference, Taipei, Taiwan.

Glassner, A., Weinstock, M., & Neuman, Y. (2005). Pupils' evaluation and generation of evidence and explanation in argumentation. *British Journal of Educational Psychology*, *75*, 105–118.

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

Hersh, R. (2005, November). What does college teach? *Atlantic Monthly*, 140–143.

Hofer, B., & Pintrich, P. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, *67*, 88–140.

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

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

Kuhn, D. (1995). Microgenetic study of change: What has it told us? *Psychological Science*, *6*, 133–139.

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

Kuhn, D., & Franklin, S. (2006). The second decade: What develops (and how)? In W. Damon & R. Lerner (Series Eds.) & D. Kuhn & R. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language*. (6th ed., pp. 953–993). Hoboken, NJ: Wiley.

Kuhn, D., & Park, S. (2005). Epistemological understanding and the development of intellectual values. *International Journal of Educational Research*, *43*, 111–124.

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

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

Kuhn, D., & Udell, W. (2007). Coordinating own and other perspectives in argument. *Thinking and Reasoning*, *13*, 90–104.

Leitao, S. (2003). Evaluating and selecting counterarguments. *Written Communication*, *20*, 269–306.

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

Miller, P. H., & Coyle, T. R. (1999). Developmental change: Lessons from microgenesis. In E. K. Scholnick, K. Nelson, S. A. Gelman, & P. H. Miller (Eds.), *Conceptual development: Piaget's legacy* (pp. 209–239). Mahwah, NJ: Erlbaum.

Moshman, D. (2005). *Adolescent psychological development: Rationality, morality, and identity* (2nd ed.). Mahwah, NJ: Erlbaum.

Nussbaum, E. M. (2003). Appropriate appropriation: Functionality of student arguments and support requests during small-group classroom discussions. *Journal of Literacy Research*, *34*, 501–544.

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., & Kardash, C. (2005). The effects of goal instructions and text on the generation of counterarguments during writing. *Journal of Educational Psychology*, *97*, 157–169.

Nussbaum, E. M., & Sinatra, G. (2003). Argument and conceptual engagement. *Contemporary Educational Psychology*, *28*, 384–395.

Perkins, D. (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.

Schwarz, B., Newman, Y., & Biezuner, S. (2000). Two wrongs may make a right ... If they argue together. *Cognition and Instruction*, *18*, 461–494.

Siegler, R. (2006). Microgenetic studies of learning. In W. Damon & R. Lerner (Series Eds.) & D. Kuhn & R. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (6th ed., pp. 464–510). Wiley.

Smetana, J., Killen, M., & Turiel, E. (1991). Children's reasoning about interpersonal and moral conflicts. *Child Development*, *62*, 629–644.

Stanovich, K. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.

Udell, W. (2007). Enhancing adolescent girls' argument skills in reasoning about personal and non-personal decisions. *Cognitive Development*, *22*, 341–352.

Voss, J. (Ed.). (2001). Argumentation in psychology. *Discourse Processes*, *32*(2 & 3).

Voss, J., & Means, M. (1991). Learning to reason via instruction in argumentation. *Learning and Instruction*, *1*, 337–350.

Wainryb, C., Shaw, L., Laupa, M., & Smith, K. (2001). Children's, adolescents', and young adults' thinking about different types of disagreements. *Developmental Psychology*, *37*, 373–386.

Walton, D. N. (1989). Dialogue theory for critical thinking. *Argumentation*, *3*, 169–184.

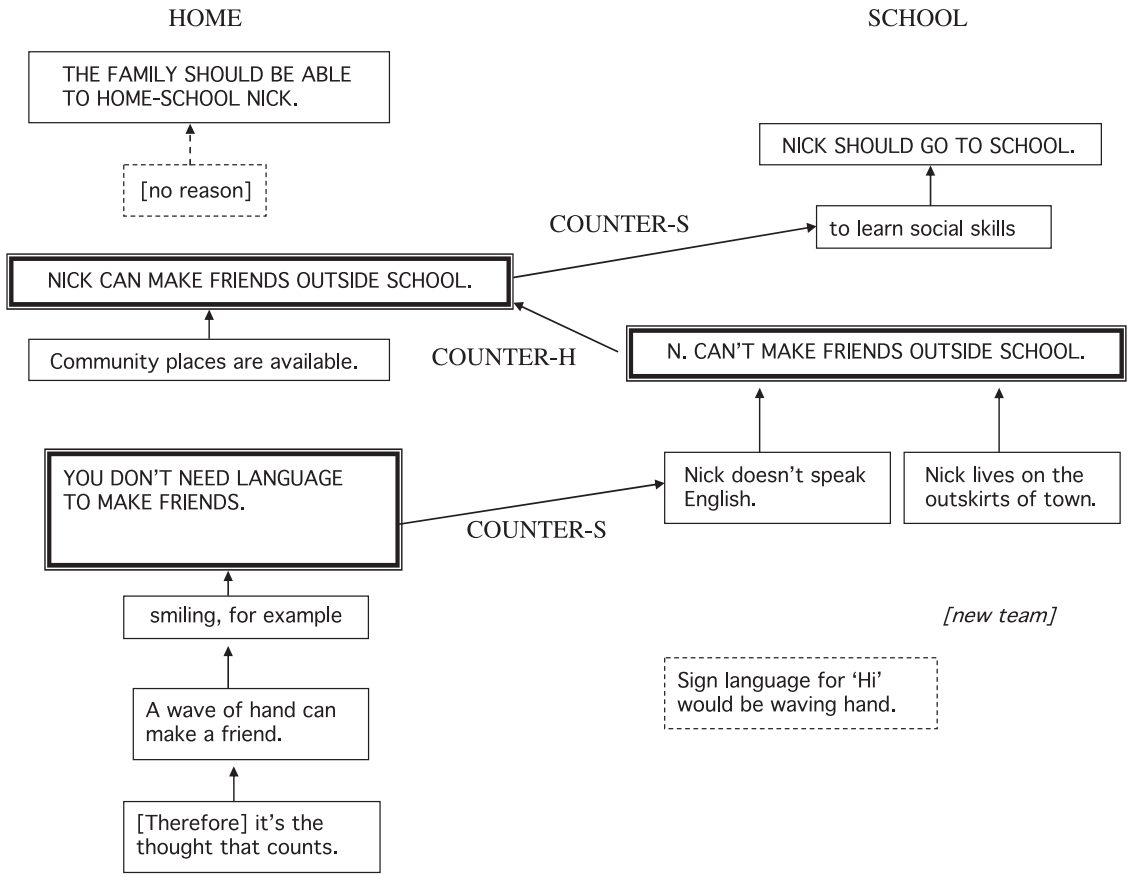
Weinstock, M., Neuman, Y., & Tabak, I. (2004). Missing the point or missing the norms? Epistemological norms as predictors of students' ability to identify fallacious arguments. *Contemporary Educational Psychology*, *29*, 77–94.

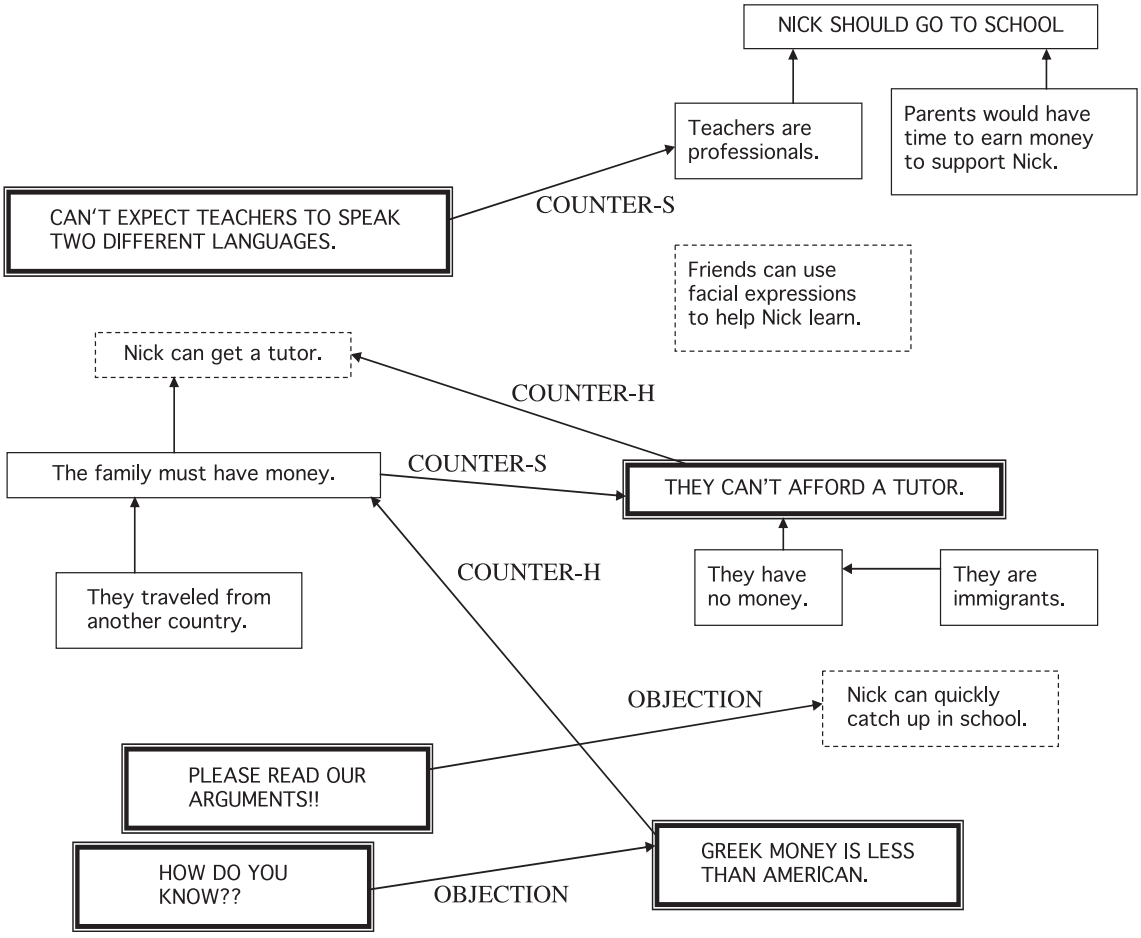
Wiley, J., & Voss, J. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology*, *91*, 301–311.

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

Appendix

Group Argument Map Produced at Homeschool Showdown




NICK SHOULD GO TO SCHOOL
Teachers are professionals.
Parents would have time to earn money to support Nick.
CAN'T EXPECT TEACHERS TO SPEAK TWO DIFFERENT LANGUAGES.
COUNTER-S
Friends can use facial expressions to help Nick learn.
Nick can get a tutor.
COUNTER-H
The family must have money.
COUNTER-S
THEY CAN'T AFFORD A TUTOR.
COUNTER-H
They traveled from another country.
They have no money.
They are immigrants.
OBJECTION
Nick can quickly catch up in school.
PLEASE READ OUR ARGUMENTS!!
HOW DO YOU KNOW??
OBJECTION
GREEK MONEY IS LESS THAN AMERICAN.