



The Development of Argument Skills

Author(s): Deanna Kuhn and Wadiya Udell

Source: *Child Development*, Vol. 74, No. 5 (Sep. - Oct., 2003), pp. 1245-1260

Published by: [Blackwell Publishing](#) on behalf of the [Society for Research in Child Development](#)

Stable URL: <http://www.jstor.org/stable/3696176>

Accessed: 16/09/2011 14:54

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Blackwell Publishing and Society for Research in Child Development are collaborating with JSTOR to digitize, preserve and extend access to Child Development.

<http://www.jstor.org>

The Development of Argument Skills

Deanna Kuhn and Wadiya Udell

This work sought to obtain experimental evidence to corroborate cross-sectional patterns of development in argument skills and to evaluate the effectiveness of an intervention designed to foster development of these skills in academically at-risk 13- to 14-year-olds. Students participated in 16 sessions of a collaborative, goal-based activity providing dense exercise of argumentative thinking. One condition included peer dialogues; another did not. The former was the more effective, although both groups progressed. Participants showed increased frequency of usage of powerful argumentative discourse strategies, such as counterargument, and decreased frequency of less effective strategies. Quality of individual arguments (for or against a claim) also improved, supporting the existence of a close relation between these two kinds of argument skills.

Educators seeking to develop thinking skills would likely consider their efforts largely successful if students became proficient in advancing, critiquing, and defending claims in reasoned discussion with peers. Yet the psychology literature has contributed little to our understanding of how such skills develop. The study of reasoning by cognitive psychologists has been limited almost entirely to solitary problem solving. Only in recent years has collaborative cognition become an object of investigation, and only a small portion of this work has been devoted to argument rather than problem solving. As a result, there exists a good deal of theoretical literature on argument and argumentation (see van Eemeren, Grootendorst, & Henkemans, 1996, for a sample), but relatively little empirical evidence has been available regarding argument skills, despite their considerable educational, as well as theoretical, significance (Yeh, 2002).

The terms *argument* and *argumentation* reflect the two senses in which the term *argument* is used, as both product and process. An individual constructs an argument to support a claim. The dialogic process in which two or more people engage in debate of opposing claims can be referred to as argumentation or argumentative discourse to distinguish it from argument as product. Nonetheless, implicit in argument as product is the advancement of a claim in a framework of evidence and counterclaims that is characteristic of argumentative discourse, and the two

kinds of argument are intricately related (Billig, 1987; Kuhn, 1991). Most of the empirical research on argument has been devoted to argument as product. Recently, however, this picture has begun to change, reflected in a landmark special issue of the journal *Discourse Processes* (Voss, 2001) that contains articles on argumentative discourse and its development.

Our primary concern in the present work was with argumentative discourse, although we also examined the individual arguments produced by participants before and after engaging in an argumentative discourse intervention. A major purpose of the present work was to confirm that the same developmental trends in discourse patterns identified in earlier cross-sectional research (Felton & Kuhn, 2001) are observed when change is induced experimentally. This is a critical step in establishing evidence in support of a developmental sequence (Kuhn, 1995a; Siegler & Svetina, 2002).

Observations of experimentally induced change are also critical to an understanding of mechanisms of change and to a conceptualization of the intersection between learning and development (Kuhn, 1995a, 1995b; Siegler, 2000; Siegler & Crowley, 1991). The position we adopt here and elsewhere (Kuhn, 2001a; Kuhn, Shaw, & Felton, 1997) is that extended exercise of thinking and reasoning skills in a cognitively rich environment can serve as a sufficient condition for their development. A design like the one used in the present study holds the potential to support this claim and to provide evidence regarding the form such development takes.

The population in which the present study was conducted should be noted. We chose to work with inner-city minority students in two low-performing

Deanna Kuhn and Wadiya Udell, Teachers College, Columbia University.

We are grateful to Anita Krishnan, Jane Monroe, and Jim Corter for statistical consultation.

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

public middle schools in New York City (one of which was closed because of poor performance shortly after we completed our study). All students attending these schools are regarded as at risk for academic failure. By eighth grade (the grade level of our participants), the large majority have become disengaged and disinterested in academic pursuits. Teachers, similarly, are largely discouraged by the formidable challenge they face and are preoccupied with disciplinary issues. Academic work focuses on the basic skills assessed in mandated standardized tests. In addition to the obvious practical importance of doing so, one of our objectives in choosing this difficult-to-work-with population was to establish that the skills that we hoped to see develop among these students were not ones emphasized in their schoolwork and hence possibly in the process of developing anyway, a hypothesis that might be advanced if students came from an academically advantaged, high-performing school.

What do we already know about the argument skills of children and adolescents? Even young children show some competence in producing arguments in support of a claim (Anderson, Chinn, Chang, Waggoner, & Yi, 1997; Clark & Delia, 1976; Eisenberg & Garvey, 1981; Orsolini, 1993; Stein & Miller, 1993) and in understanding the structure of an argument (Chambliss & Murphy, 2002). Educational studies have documented that constructing arguments (Voss & Wiley, 1997; Wiley & Voss, 1999; Zohar & Nemet, 2002) and engaging in argumentative discussion (Mason, 1998, 2001) enhance conceptual understanding of subject matter in school-age children, as well as college students.

Nonetheless, serious weaknesses have been observed in the arguments of adolescents and young adults. They are unlikely to construct two-sided arguments or to distinguish evidence and explanation in support of their claims (Brem & Rips, 2000; Kuhn, 1991, 2001b; Kuhn et al., 1997; Perkins, 1985; Voss & Means, 1991). College students show some skill in evaluating arguments (Rips, 2002) but also significant weaknesses, especially in susceptibility to belief bias (Klaczynski, 2000). The evidence available regarding argumentative skills in classroom discourse is consonant with this picture. Pontecorvo and Giradet (1993) reported that the large majority (81%) of utterances by small groups of 9-year-olds asked to reach agreement about a historical claim were devoted to espousing their own claims and justifications of them.

Instructional units devoted to construction of arguments have been found productive in enhancing the quality of arguments supporting a claim

(Hidi, Berndorff, & Ainley, 2002; Knudson, 1992). Consistent with the theoretical perspective indicated earlier, however, Kuhn et al. (1997) and Lao and Kuhn (2002) have shown that extended engagement in argumentative discourse, in the absence of any additional instruction, is a sufficient condition for enhancement of the quality of arguments produced by individuals following discourse. Studies by Anderson and colleagues (Anderson et al., 1997; Anderson, Chinn, Waggoner, & Nguyen, 1998; Anderson et al., 2001; Chinn & Anderson, 1988; Reznitskaya et al., 2001), which we discuss in more detail later, also support this conclusion.

We focused our present work on skill in argumentative discourse itself. Developing a coding scheme for analyzing such discourse is a formidable task (for examples of existing efforts, see Berkowitz & Gibbs, 1983; Felton & Kuhn, 2001; Keefer, Zeitz, & Resnick, 2000; Rips, Brem, & Bailenson, 1999). As a result, much less research has been devoted to argumentative discourse than to arguments that are the cognitive constructions of individuals.

What sorts of discourse skills might we expect to observe among young adolescents, and in what directions would we hope to see these skills develop? 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. Drawing on Walton's analysis, Felton and Kuhn (2001) identified two potential forms of development in argumentative discourse skills: enhanced understanding of discourse goals and application of effective strategies to meet these goals. These two forms of development can be predicted to reinforce one another. Progress in use of discourse strategies is propelled by a better understanding of discourse goals. At the same time, exercise of these strategies in discourse promotes more refined understanding of the goals of argumentative discourse.

To examine development in argumentative discourse skills, Felton and Kuhn (2001) conducted a cross-sectional comparison of the dialogues of young teens and community college young adults arguing about capital punishment (CP). The results revealed striking differences between the two groups. Teens' discourse focused largely on the arguments supporting their own position, at the expense of addressing the arguments of their opponents. Adults, in contrast, in addition to advancing their own arguments, were more likely to address the opponent's argument, most often through counterargument. In

undertaking to undermine their opponent's argument, as well as advance their own argument, adults' dialogues thus came closer to achieving the dual goals of argumentative discourse. These appear to be skills that need to develop during childhood and adolescent years. Deep-level processing of the opponent's argument, in addition to articulating one's own argument and negotiating the mechanics of discourse, may represent cognitive overload for the novice arguer.

The cross-sectional work by Felton and Kuhn (2001) allowed us to make the prediction that initial argumentative dialogues of the eighth graders in the present study would consist predominantly of what we term *exposition*, that is, articulation and clarification of one's own position and perspective. A relatively small proportion, in contrast, we predicted, would consist of *challenges* that address the partner's claims and seek to identify weaknesses in them, reflecting understanding of Walton's (1989) second goal of argumentation. The goal of the activity undertaken in this study was to effect a shift in the direction of an increasing proportion of dialogue devoted to challenge and a decreasing proportion devoted to exposition.

A further purpose of the present study was to examine the role of argumentative discourse in the change we observe in argument skills. To fulfill this objective, we developed an intervention designed to scaffold components of argumentative discourse skill (listed in Table 1). The intervention consisted of two phases. In the first phase, teams of participants collaborated to develop their own argument to justify their chosen (pro or con) position. In the second phase, teams engaged in various ways with the opposing team. The design of the study was such

that participants randomly assigned to one group engaged in the entire intervention, whereas participants randomly assigned to a comparison group engaged only in the first phase. If engagement in discourse with an opposing side is important to progress, advancements in discourse skill should be concentrated in the first group.

Three aspects of argument skill were assessed: (a) the quantity of different reasons a participant has available in his or her knowledge base on the topic as potential components of an argument, (b) the quality of argument produced by the individual (based on the assessment scheme devised by Kuhn et al., 1997), and (c) the quality of argumentative discourse a participant produces in dialogues with a peer (based on the assessment scheme developed by Felton & Kuhn, 2001). Each of these dimensions was assessed at the outset of the study and again following intervention.

Method

Participants

Participants were 34 academically at-risk eighth-grade students attending two low-performing, inner-city public middle schools in New York City. The two schools were comparable and served similar populations. The percentages of the student body meeting state standards in math were 17% in one school and 3% in the other. Percentages meeting state standards in reading were 28% and 10% in the two schools, respectively. Percentages of students qualifying for free lunch were 91% in one school and 72% in the other. Of the 34 participants, 14 were

Table 1
Summary of Activities and Cognitive Goals

Activity	Goal
Phase I	
Generating reasons	Reasons underlie opinions. Different reasons may underlie the same opinion.
Elaborating reasons	Good reasons support opinions.
Supporting reasons with evidence	Evidence can strengthen reasons.
Evaluating reasons	Some reasons are better than others.
Developing reasons into an argument	Reasons connect to one another and are building blocks of argument.
Phase II	
Examining and evaluating opposing side's reasons	Opponents have reasons, too.
Generating counterarguments to others' reasons	Opposing reasons can be countered. "We can fight this."
Generating rebuttals to others' counter arguments	Counters to reasons can be rebutted. "We have a comeback."
Contemplating mixed evidence	Evidence can be used to support different claims.
Conducting and evaluating two-sided arguments.	Some arguments are stronger than others.

Hispanic, 19 were African-American, and 1 was Ethiopian. Of the 34, 11 were male and 23 were female. A total of 21 students participated in the experimental condition and the remaining 13 participated in the comparison condition. Five additional students (3 in the experimental group and 2 in the comparison group) began but did not complete the study because of chronic absenteeism or suspension.

Performance Measures

Individual assessment of arguments. The study began and concluded with individual assessment of a student's opinion on CP (on a 13-point scale developed by Kuhn & Lao, 1996) and the student's reasons in support of that opinion. When a probe ("Anything else?") elicited no more reasons, the student in addition was asked to indicate all of the reasons he or she could think of that someone might give to support pro or con opinions on CP. The number of unique reasons given in response to the initial elicitation of reasons and to the "all reasons" question was taken as the measure of the number of CP reasons available in the student's knowledge base.

Responses to the initial question provided the basis for coding of the participant's quality of argument based on the scheme developed by Kuhn et al. (1997) and shown in Table 2. Arguments were coded as one-side or two-sided depending on whether they included both pro and con reasons or only one of the two. In addition, arguments were coded as to the highest level of reason shown, as well as to presence of the specific reasons listed in Table 2. Level III reasons for or against CP, the lowest level in the hierarchy (see Table 2), were regarded as having little or no argumentative force. Level II CP reasons were regarded as nonfunctional reasons as they did not address the functions of CP and instead focused on conditions under which it should be administered or potentially remediable defects in the way it is administered, rather than its purposes. Level I reasons addressed the purposes or functions of CP. Within Level I, a further distinction was made between reasons that address the desirability of CP within a framework of alternatives (Type A, Table 2) and those that simply offer reasons for or against CP, without consideration of its alternatives (Types B and C; see Kuhn et al., 1997, for further discussion of the coding scheme.) Interrater reliability was calculated on 70% of the responses. Percentage agreement between two raters was 81% (Cohen's kappa = .67).

Assessment of argumentative discourse. Following individual assessment, pairs of students who held

contrasting CP opinions were formed. With few exceptions (because of the attrition described earlier), participants met with the same partners in the initial and final assessments. Each pair met in a room with an adult who videotaped the dialogue for subsequent analysis. The pair was given the following instruction:

We are here for a discussion about capital punishment. Your task is to talk about the issue of capital punishment, of whether a person should be put to death for a serious crime. I'd like the two of you to have a serious discussion about the issue. You'll need to discuss the reasons each of you have for your views and find out where you agree and disagree. If you disagree, try to determine why and try to reach an agreement if you can.

The pair was informed they had 10 min to complete their dialogue. They were given a 1-minute signal after 9 min.

Dialogues were transcribed and coded according to the scheme developed by Felton and Kuhn (2001; see the Appendix for a summary of the codes). Each utterance in the dialogue was segmented and assigned one of the codes shown in the Appendix. Four raters participated in coding. Raters were trained on a randomly chosen subset of dialogues that constituted 15% of the database. Raters completed training when the percentage agreement among raters reached 84%, excluding utterances agreed on after initial disagreement and discussion. The dialogues used for training were recoded and checked for reliability. The remaining dialogues were divided equally among three of the raters. The fourth rater randomly sampled 25% of the coded dialogues and recoded them. Percentage agreement between the fourth rater and a set of randomly sampled dialogues was 90% (Cohen's kappa = .72).

Intervention

Participants were assigned to pro and con teams based on their opinions at the initial assessment. At each of the two participating schools, there was a pro and con team ranging from four to eight members. Because of the high absence rate at these schools, actual attendance on a given day tended to range from three to six members per team. A second pro and con team at one of the schools served as the comparison group; their activity ended after the first 7 (of 16) sessions had been completed (at the end of

Table 2
Coding Scheme for Individual Arguments

Pro arguments
<p>I. Functional arguments</p> <ul style="list-style-type: none"> A. Alternatives to CP are ineffective or less effective than CP <ul style="list-style-type: none"> A1. Alternatives to CP are not effective as deterrents A2. Alternatives to CP are not effective in protecting society from criminals A3. Alternatives to CP are not sufficient punishment A4. Alternatives to CP fail to rehabilitate criminals A5. Alternatives to CP are too burdensome or costly a way to serve their purpose B. CP reduces crime <ul style="list-style-type: none"> B1. CP deters people from crime B2. CP protects society from the acts of criminals C. CP is an appropriate punishment <ul style="list-style-type: none"> C1. Eye-for-eye C2. Criminals have forfeited the right to life and privileges associated with it C3. Compensates victim or victim's family <p>II. Non-functional arguments (focused on conditions that make CP justified, without consideration of its functions)</p> <ul style="list-style-type: none"> A. CP is justified only if guilt is established beyond reasonable doubt B. CP is justified only if criminal judged competent to be responsible for own actions C. CP is justified only if it is applied consistently D. CP is justified only if the crime is sufficiently grave E. CP is justified only in the case of repeated crime <p>III. Nonjustificatory arguments</p> <ul style="list-style-type: none"> A. Justification based on sentiment B. Appeal to precedent (CP has been in use for a long time) C. Appeal to majority (many or most think it's a good idea) D. Appeal to authority (without intervening argument) E. Crime exists and needs a remedy
Con arguments
<p>I. Functional arguments</p> <ul style="list-style-type: none"> A. Alternatives exist that are preferable to CP <ul style="list-style-type: none"> A1. Alternatives to CP are better as deterrents A2. Alternatives to CP are better in protecting society from criminals A3. Alternatives to CP are better punishment A4. Alternatives to CP allow rehabilitation of criminals B. CP does not reduce crime or reduce it sufficiently <ul style="list-style-type: none"> B1. CP is not effective in deterring people from crime B2. CP is not effective in protecting society from the acts of criminals C. CP is not an appropriate punishment <ul style="list-style-type: none"> C1. CP commits the same crime it is meant to punish C2. CP does not right the wrong (doesn't restore loss to victim of crime) C3. We lack the right to take life C4. We lack the right to make judgments of who should live or die C5. We lack the right to make judgments of other people's actions C6. CP violates the principle of forgiveness C7. Any killing is wrong C8. CP is violent, barbaric C9. CP wastes lives C10. CP serves no purpose C11. Enforcers of CP themselves commit crime <p>II. Nonfunctional arguments (focused on possibly remediable defects in administration of CP, without consideration of its functions)</p> <ul style="list-style-type: none"> A. CP may punish innocent people B. CP may punish people who are not responsible for their actions C. CP is not administered uniformly (may be discriminatory against certain groups) D. CP may punish people who committed crime accidentally or as victim of circumstances E. CP is not administered efficiently (e.g., may be drawn out and costly) <p>III. Nonjustificatory arguments</p> <ul style="list-style-type: none"> A. Justification based on sentiment B. Appeal to precedent (CP has not been widely used or as widely used as it once was) C. Appeal to majority (many or most are against CP) D. Appeal to authority (without intervening argument)

Note. CP = capital punishment. Source: Kuhn, Shaw, and Felton (1997).

Phase I, after the developing reasons into an argument activity; see Table 1).

Sessions were scheduled to meet twice a week for a 90-min period for a total of 16 sessions over 8 weeks. Standardized testing, school assemblies, holidays, and two school vacations, however, extended the total period in which the 16 sessions took place to most of one semester, approximately 12 weeks.

The procedure was identical for the experimental and comparison groups through the first seven sessions (through Phase I in Table 1). Thus, members of the comparison group were introduced to the same goal-based activity as members of the experimental group, and they participated with the experimental group in developing reasons into an argument, but they did not engage in argumentative discourse with peers who held a contrasting opinion (see description of activities that follows). The rationale was to have comparison participants take part in as full and engaging argument-based activity as feasible, minus the dialogic component.

At each of the sessions, each pro and con team met separately with an adult coach. The sequence of activities and the goals associated with them are summarized in Table 1. Teams progressed through this sequence at approximately the same pace, but with latitude for participants and coach to stay with an activity until they reached closure. On average, about one and one half sessions were devoted to each activity (with its associated goal) listed in Table 1. Here we briefly describe the kinds of activities in which the teams engaged. In most of the activities described, participants broke into two or three pairs or triads to work independently and then reassembled to discuss their work as a whole group.

Phase I

Generating reasons. At the initial meeting it was explained that the team would be working together for several weeks to prepare for a "showdown" in which they would debate the topic with a team who held the opposing view. An outing, it was explained, would take place to celebrate the winners' victory. The day's activity began with participants asked to remember the reasons they gave for their opinion in the initial interview and to think of which were the most important reasons. Large index cards were distributed and participants wrote their most important reason on a card. The coach provided a rationale for each activity in which the team engaged; for this activity, she said, "The first step is to be clear about *why* we think CP is good. We saw

in talking with each of you that people can have different reasons for thinking it's good, so we need to get these reasons out on the table and decide what we think of them." (The pro position is assumed for this description.) Participants were encouraged to interpret reasons ("What does this one mean?" "Is there a different way to say this?") and eliminate duplicates. Participants then generated and recorded as many additional reasons as they could, and the reflection and elimination of duplicates was repeated. As a homework assignment, they were asked to bring in for the next session the reasons of three same-side people they queried, and at the next session these reasons were dealt with in the same way. (The coach added at most one reason to this discussion if she identified any major reasons the group had not generated.)

Elaborating reasons. Participants were asked to reflect on whether these were good reasons, leading to a discussion of what makes a reason a good one, and to revise the wording of some reasons. The criterion of how a person with an opposing view would react to the reason was introduced. Possible ways to strengthen reasons were considered. The coach introduced the distinction between why reasons, justifying why CP is a good idea (Level I in Table 2), and when reasons, specifying conditions for its use (Level II in Table 2). Each participant chose ownership of at least one reason for which he or she would be responsible.

Supporting reasons with evidence. The concept of evidence as strengthening a reason was made explicit. Students discussed types of evidence and were provided several newspaper articles with stories or statistics that supported their position. Articles were read in pairs and reported to the group, who decided how the information might strengthen any of their reasons. A summary of supporting evidence was recorded with each reason.

Evaluating reasons. In pairs or triads, participants were given duplicate sets of the team's reason cards and asked to sort them into three categories (best, good, and okay), discussing placement until agreement was reached. This activity was repeated with new subgroups and finally as a whole group. Reasons remained in the top category only if participants could justify (with reasons for reasons) to the group's satisfaction why it belonged there.

Developing reasons into an argument. Initially in pairs and then together, the team worked to construct and format on poster board an argument based on their set of strong reasons. This activity required discussion of which reasons to use, the

relations between reasons and the possibility of connecting words or phrases, the order of presentation, and the inclusion of examples and evidence. After practice presentations, the group chose one member to present the argument, which was recorded on video for subsequent analysis and critique by the group. How the presentation could be used at the showdown was discussed.

Phase II

Examining and evaluating opposing side's reasons. The question was raised of what the opposing team had been doing during this time and what reasons they would have, leading to the recognition that it would be useful to know what their reasons were and to an effort to anticipate them. The coach led participants to the recognition that it would be necessary to counter these reasons ("What is the value of knowing the other side's reasons? What will we want to do with them?"). The coach then disclosed that she had made a deal with the opposing coach to exchange reasons and distributed duplicate sets of the opposing team's reasons. After some animated examination and discussion, participants were divided into subgroups to evaluate the opposing side's reasons by sorting them into strong, middle, and weak categories. The format for this activity was identical to that for the earlier activity of evaluating their own side's reasons.

Generating counterarguments to others' reasons. The generation of counterarguments, which began spontaneously in the preceding activity, was formalized. Participants were divided into pairs and were given colored *counter* cards on which to record counterarguments to each reason by the other side. Pairs were formed again and the whole group then reassembled to deliberate which candidate was the strongest counter to the other side's reason. When agreement was reached, that counter was attached to the reason card and one member of the team took ownership of it.

Generating rebuttals to others' counterarguments. The team's reason cards were returned to them, with counterargument cards produced by the opposing team attached. In pairs, team members debated how to respond—whether they should strengthen their reason to avoid the criticism, rebut the criticism, or drop the reason. Where appropriate, rebuttals were generated. The criterion of reducing the strength of the counterargument was identified. After different pairings, the whole group met to agree on a resolution, decide on the best rebuttals, attach rebuttal cards to the reason-counterargument

card pairs, and assume ownership of each of the card sets for the showdown.

Contemplating mixed evidence. The coach offered the team some additional evidence material she had identified. Some pieces were the news articles they had seen initial paragraphs of earlier but now with the later paragraphs (representing the opposing perspective) restored. All of the articles included material that could be drawn on by either side. The coach encouraged the team to consider how the opposing team would use evidence as well as how the evidence could strengthen their position.

Conducting and evaluating two-sided arguments. As preparation for the showdown, the coach played the opposing side in reviewing argument-counterargument-rebuttal sequences. As final preparation, opposing team members came to the room for practice dialogues, which remaining team members critiqued.

Showdown. Two opposing teams met and reviewed the rules for the showdown. Each team chose who would speak for the team and when a new speaker would be substituted, with the provision that no person would speak for more than 3 min and every person would speak at least once. At the request of any team member, the two teams broke into 1-min huddles to confer. The debate continued for approximately 20 min, after which the judge announced that it would be necessary to conduct some tie-breaking dialogues between pro-con pairs, which would be conducted later that week.

Results

We begin with analysis of the discourse results, then turn to the effect of the activity on the CP arguments given by individuals.

Changes in Argumentive Process

Our major hypotheses, as noted earlier, were a decrease in discourse devoted to exposition and an increase in discourse devoted to challenge. In the scheme developed by Felton and Kuhn (2001; see the Appendix), three major categories of utterances (as well as several infrequently used minor utterances) fall under the heading of challenge. One is simple disagreement (with what the partner has said). The two more advanced types that fall under this heading are Counter-A and Counter-C. Counter-A is a disagreement accompanied by an alternate argument. Counter-C is disagreement accompanied by a critique of the partner's utterance.

Table 3
Mean Numbers (and Standard Deviations) of Utterances per Participant per Dialogue

	Mean utterances	
	Initial	Final
Experimental	7.23 (5.06)	29.58 (17.56)
Comparison	7.15 (4.25)	11.11 (9.83)

The second major hypothesis was a decline in the proportion of discourse devoted to exposition. Two major categories in the Felton and Kuhn (2001) scheme that fulfill this definition are Clarify and Add. The Clarify category is limited to statements that elaborate one's own position. The Add category consists of statements that add something (e.g., an example or an elaboration) to what the partner has said, but in a conversational manner that does not advance the goals of argumentation. Both of these types we hypothesized would decrease in frequency following intervention.

Because of a recording error, a dialogue of 2 experimental participants was lost, and the analysis of dialogues is based on a sample of 32 participants. As seen in Table 3, final dialogues were substantially longer than initial dialogues among members of the experimental group but showed negligible change in length among comparison participants. Each of the utterances an individual made in a dialogue was assigned to one of the codes in the Appendix. Note that this is a functional coding system of dialogic argument (in contrast to the substantive coding system in Table 2). An utterance is assigned a code based on its function, relative to the immediately preceding utterance of the opposing speaker, rather than its content (see the Appendix).

To allow comparison of the two groups with respect to the types of utterances contained in their dialogues, the frequencies of occurrence of each utterance type for each participant were converted to percentages (of the total number of utterances made by the individual in that dialogue). Omitted from the analysis were utterances that were off topic or were metastatements about the dialogues (rather than contributions to them). Metastatements ranged from fairly high level (e.g., "That's what I'm saying") to low level (e.g., "What are we supposed to be doing?") and averaged from less than 1% to 10% (the latter for comparison group participants at the final assessment).

Summarized in Table 4 are the results of statistical analysis of proportions of usage in posttest dialo-

Table 4
Initial and Final Mean Percentage Use (and Standard Deviation) of Discourse Types by Condition

	Percentage use	
	Initial	Final
Exposition type (predicted to decrease)		
Clarify ^a		
Experimental	37.2% (30.4)	18.5% (11.3)
Comparison	44.3% (31.0)	32.4% (22.1)
Add ^b		
Experimental	7.3% (14.3)	1.7% (3.4)
Comparison	1.5% (3.8)	7.5% (10.9)
Challenge type (predicted to increase)		
Counterargument: critique ^c		
Experimental	5.3% (7.9)	30.6% (12.2)
Comparison	11.4% (16.9)	21.4% (24.1)
Counterargument: alternative ^d		
Experimental	7.6% (13.7)	13.6% (8.5)
Comparison	5.3% (8.0)	1.2% (2.8)
Disagree ^e		
Experimental	1.4% (4.2)	6.3% (7.6)
Comparison	7.7% (2.8)	1.3% (3.3)

^aSignificant effect of group at posttest, $F(1, 29) = 8.91, p = .006$.

^bSignificant effect of group at posttest, $F(1, 29) = 5.82, p = .02$.

^cSignificant effect of group at posttest, $F(1, 29) = 4.21, p = .049$.

^dSignificant effect of group at posttest, $F(1, 29) = 21.07, p = .001$.

^eMarginally significant effect of group at posttest, $F(1, 29) = 3.22, p = .08$.

gues of the two utterance types classified as exposition and the three classified as challenge, for experimental and comparison groups. An arcsin transformation was used to normalize these proportions. Individuals' pretest proportions were used as a covariate and the data were subjected to a multivariate analysis of covariance (MANCOVA). An overall multivariate effect for group (experimental vs. comparison) was significant, $F(5, 25) = 6.0, p = .001$, holding pretest scores constant, pretest $F(5, 25) = 3.55, p = .015$, allowing us to examine the effect of group with respect to the specific predictions made for individual dependent variables. These results are shown in Table 4. The prediction of decline for the two exposition types (reflected in the group differences at posttest) was confirmed. In the more prevalent of the two types—clarify—both groups show decline, but the experimental group shows steeper decline, producing the group difference at posttest. The prediction of increase for the three challenge types (similarly reflected in group differences at posttest) was also confirmed with the qualification that the group difference for the least advanced challenge type—disagree—only approached significance.

Table 5
Examples of Counter-C and Counter-A Discourse Types

Assertion
AL: I think I'm in favor of capital punishment because if you did a crime and took away many lives like the Unibomber, I think you should pay for that crime....

Counter-A
DA: Yeah, I thought the same thing at first but then I thought that it was wrong because you are taking a life away and I think that's plain out wrong because the only person who could take a life away is God. God gives you the life, God takes it away....

Counter-C
AL: But you don't agree that the Unibomber, he took lots of life away, you don't think that he should pay for that?

Assertion
GL: Somebody can be let out for good behavior and they might come out and do the same thing again.

Counter-C
TW: But it takes awhile, though. It takes years to come.

Counter-C (Rebuttal)
GL: Okay, but they will come out eventually and try to get revenge. Then what's going to happen to somebody else, maybe from your family?

Counter-A
TW: But they chose to do it, though. Nobody made them do it.

Note. These are excerpts from actual dialogues. Coding of each statement is based on its functional relation to the partner's immediately preceding statement. The two excerpts presented are representative of the quality of dialog observed in this sample, with the first excerpt toward the high end and second excerpt toward the low end but still fulfilling the criteria for the codes assigned.

Of the three challenge types, Counter-C was the one most likely to represent a newly acquired competence (compared with simply disagreeing or expressing an opposing reason). It is therefore relevant to examine how common it was for an individual to exhibit this competence following the intervention. As reflected in Table 4, both groups show increased usage. Among experimental participants, 8 of 19 exhibited at least one Counter-C at the initial assessment. At the final assessment, all 19 did so. Among comparison participants, 5 of 13 exhibited at least one Counter-C at the initial assessment. At the final assessment, an additional 5 participants had done so, for a total of 10. Thus, the comparison condition was sufficient to invoke the appearance of Counter-C in only some of the participants who did not already display it. The experimental condition, in contrast, was always successful in doing so. Examples of Counter-C and Counter-A excerpts from the posttest dialogues are shown in Table 5.

Counter-C sequences. When an opponent's counterargument is followed directly by another counterargument introduced by the original speaker, the second counterargument functions as a rebuttal that removes or reduces the force of the initial counterargument. Because of the theoretical significance of rebuttals, we examined all dialogues for sequences of counter utterances, defined as a Counter-C or Counter-A followed directly by a Counter-C.

At the initial assessment, only two comparison and three experimental participants exhibited any

Table 6
Frequency (and Standard Deviation) of the Rebuttal Sequence

	Mean frequency per participant	
	Initial	Final
Experimental	0.16 (0.50)	3.79 (1.75)
Comparison	0.23 (0.44)	0.46 (0.66)

rebuttals. At the final assessment, 5 (of 13) comparison participants exhibited rebuttals, as did all 19 experimental participants. Thus, although the comparison intervention was sufficient to promote at least some counterargument usage among most (10 of 13) comparison and all experimental participants, only the experimental intervention was powerful enough to foster rebuttal skill in all participants.

The absolute number of Rebuttal sequences was calculated for each participant in each dialogue. Means, across all participants, appear in Table 6. Both experimental and comparison groups increased use of rebuttals from initial to final assessment, but the experimental group's increase was greater, to an average of almost four rebuttals per dialogue. The main effect of time was significant, $F(1, 30) = 41.76$, $p < .001$, as was the effect of condition, $F(1, 30) = 47.44$, $p < .001$, and their interaction, $F(1, 30) = 32.37$, $p < .001$.

Remaining codes. Of the remaining discourse types (see the Appendix), all but three showed only

minimal usage (averages of less than 5% across participants). Of the three discourse types showing greater than 5% usage, two showed no change over time and one showed a decrease. The two types showing no change were Case-? and Agree. Use of the Case-? type ranged from 3% to 6%, with the experimental group decreasing usage more than the comparison group over time. Use of the Agree type ranged from 6% to 11%, with the experimental group decreasing and the comparison group increasing. Statistical comparisons were nonsignificant for these types.

A final discourse type, Position-?, decreased in frequency from initial to final assessment for both groups (from 9.7% to 0% for the experimental group and from 11.1% to 1.6% for the comparison group). It consisted of a request for the partner to state his or her position on an issue. This decline is a procedural result of the fact that almost all participants began their initial dialogues with an attempt to establish one another's positions on the topic (e.g., "So, are you for or against it?"). By the time of the final assessment, all participants had become very familiar with the topic, discussed it extensively, and were aware of one another's positions. Hence, the final dialogues tended to omit this opening exchange.

Changes in Argumentive Products

We turn now to how the intervention activity affected the individual arguments offered by participants.

Changes in knowledge base. At the initial individual assessment, before intervention, experimental participants offered an average of 3.30 reasons from the list in Table 2, whereas comparison participants gave an average of 2.31 reasons. These means include responses to both the initial elicitation of reasons and the "all reasons" question. Omitted from this analysis is 1 experimental participant for whom "all reasons" responses are missing. Because of this chance difference between the two groups (because assignment to groups was randomized), comparisons focus on changes from initial to final assessment. At the final assessment, 60% of experimental participants gave an increased number of reasons, 25% gave the same number, and 15% gave a decreased number. Comparable percentages for comparison participants were 54% increasing, 15% unchanged, and 31% decreasing. These percentages can be compared with the change observed in the simple (pretest-posttest only) control condition conducted by Kuhn et al. (1997), whose participants were from the same population. In that condition,

positive change was less frequent; only 27% gave an increased number of reasons from initial to final assessment and the modal pattern was no change, in contrast to the modal pattern of increased reasons in the present study. Hence, both conditions in the present study elevated performance beyond a baseline, no-intervention level.

By the final assessment, the mean number of reasons given by experimental participants in the present study had increased to 4.45, a mean increase of 1.15 reasons, whereas the comparison participants' mean had increased to 2.54, a mean increase of .23 reasons. The increases for the two groups were not significantly different. Closer inspection of these data, however, showed that although comparison participants offered similar sets of reasons at initial and final assessments, experimental participants offered several new reasons at the final assessment and excluded some of the reasons they had given at the initial assessment. Therefore, the total number of different reasons offered (at either of the assessments) was computed for each participant. This analysis showed experimental participants produced a larger number of reasons overall than did comparison participants. The mean number of reasons produced by experimental participants was 6.70 and the mean number produced by comparison participants was 4.38, a significant difference, $t(31) = 3.23$, $p = .003$. Further examination of the reason changes on the part of experimental participants indicated that the reasons they dropped from initial to final assessment were more likely to be lower level reasons (Levels II and III in Table 2), whereas new reasons added at the second assessment were more likely to be higher level reasons (Level Type I). Experimental participants dropped an average of 1.95 reasons from initial to final assessments and added an average of 3.30 reasons. Of the dropped reasons, 56% were lower level, whereas only 19% of added reasons were lower level. Consistent with this contrast, 29% of added reasons were of the highest, comparative (Level IA) level, whereas only 8% of dropped reasons were of this level.

Changes in knowledge of reasons, then, lie not in the number of reasons offered on each occasion but in the kinds of reasons offered, with experimental participants likely to add higher quality reasons and drop lower quality reasons. As a result, they offered a larger total number of reasons (over both occasions) than did the comparison group. This difference leads to the question of whether experimental participants improved in the quality of their arguments.

Table 7
Frequency of Types of Change in Individual Arguments

	Change type				Total
	Positive	Mixed	None	Negative	
Experimental	15	1	3	2	21
Comparison	5	2	4	2	13

Changes in argument quality. Table 7 summarizes changes in argument quality shown by participants from initial to final assessment. Possible types of change follow those identified by Kuhn et al. (1997). Four types of change occurred in the present study: (a) change from an unsupported opinion to an opinion supported by reasons, (b) change from an argument containing only lower level (Level II or III reasons, Table 2) to one containing higher level (Level I) reasons, (c) change from a noncomparative Level I argument (Level IB or IC reasons) to a comparative Level I argument (one that includes some Level IA reasons), and (d) change from a one-sided to a two-sided argument (both pro and con reasons are represented). Arguments showing one or more of these changes were classified as showing positive change. Arguments showing one or more such changes in the reverse directions were classified as showing negative change. Arguments showing a pattern of some positive and some negative changes were classified as showing mixed change.

As seen in Table 7, 71% of experimental participants and 38% of comparison participants fell into the positive change category (a difference just below statistical significance). The most frequent type of positive change was the shift from a one-sided to a two-sided argument, a shift exhibited by 8 of the 15

experimental participants showing positive change. An example of one participant's initial and final arguments that reflects this change is presented in Table 8. The three other types of positive change noted earlier were also shown by multiple participants, with some participants showing more than one type.

In sum, the comparison group showed a tendency to increase slightly the number of reasons they had available, and a substantial minority showed improvement in the quality of their arguments from initial to final assessment. Only the experimental participants, however, who engaged in the full intervention activity showed a consistent influence of this activity on the individual arguments they offered at the final assessment.

Discussion

The present work supports earlier findings that argument skills develop and that engagement in an argumentative discourse activity enhances that development (Felton & Kuhn, 2001; Kuhn et al., 1997). The contribution of the present work is to show that such advancement can be observed not only in the arguments that an individual constructs in support of a claim but also in the quality of argumentative discourse generated in peer dialogues.

Most participants in the experimental condition showed improvement in their individual arguments for or against CP, as well as advances in their argumentative discourse skills. By comparison, in earlier work (Kuhn et al., 1997) less than half of participants who engaged in a series of peer dialogues showed such improvement. Thus, the goal-based aspect of the present intervention has some advantage over discourse alone in improving individual arguments as well as discourse skills. In

Table 8
Illustration of One Participant's Progression From a One-Sided to a Two-Sided Argument

Initial assessment:

If someone did something wrong, they should be subject to capital punishment. (Why is that?) Because for instance if they kill someone, maybe the same thing is due to them. (Any other reason?) Well, I feel that people should pay if they did something wrong.

Final assessment:

If someone goes out and kills another person they should receive a justified punishment, an equal punishment. So that if they killed someone then they should receive the same thing. But I can also see how other people can have a different opinion because not everyone thinks the same and they may feel that it's wrong to kill another person, that people deserve a second chance. But personally I feel that if you have enough nerve to go out and kill somebody else, well then you just deserve to be killed as well. (Okay, anything else?) Well, one of the reasons why I have this opinion is that I've seen where facts have shown that capital punishment has reduced crime. And I always think that less crime will make a better life for everyone.

addition to the goal-based nature of the intervention, which engaged students' interest and supported their involvement, a significant feature of the intervention, we believe, is the external representation of ideas (through the card procedure and discourse) that made the argument-counterargument-rebuttal structure explicit. By engaging as well in scaffolded argumentive discourse, students gain an enhanced overall sense of what an argument consists of, what Reznitskaya et al. (2001) called an argument schema.

It is notable that progress in these respects is observable among inner-city, severely disadvantaged young adolescents, such as those who participated in the present study. The school and personal lives of these young adolescents have provided few opportunities for the practice of sustained argumentive discourse, and yet the skills we have identified show significant development when students engage in a collaborative, goal-based activity involving these skills. In current work, Kuhn and DeFuccio (2002) have obtained similar results working with an even more disadvantaged population: 16- to 18-year-old inmates in a juvenile detention center.

Which aspects of our intervention are responsible for the changes in skill levels observed in this study? Our comparison condition established that it is not involvement in an engaging, argument-focused activity itself. If it were, participants in the comparison condition would have shown greater gain. Because they invested less time overall in the activity than did experimental participants, the possibility cannot be ruled out that comparison participants might have benefited to a greater degree if we had simply prolonged or repeated the Phase I activities, so that their time investment equaled that of the experimental group. This possibility is of practical importance and deserves to be followed up.

The data from the present study, however, implicate discourse activity itself as an important element of the experience needed to optimize development of argument skills. The exercise in argumentive discourse provided in the experimental condition, rather than simply time devoted to topic-related activity, appears necessary for greatest skill development. Supporting this interpretation is the fact that comparison participants did not increase their usage of the Counter-A discourse strategy. This is significant because that strategy entails countering the opposing claim with a new same-side reason (one that does not directly address the opponent's claim). Comparison participants had as much experience as did experimental participants in develop-

ing their own side's reasons during Phase I of the intervention and hence presumably had them available to an equal extent. Yet they did not use these reasons in the service of discourse goals to the extent that was seen among experimental participants.

Despite the apparent importance of discourse, the activity in which we engaged the experimental group in the present study was more effective in developing argument skills than the discourse practice alone in the Kuhn et al. (1997) study. Teens from the same population studied here who engaged in a succession of dialogues with peers over several weeks (Felton, *in press*; Kuhn et al., 1997) less frequently showed improvement in their individual arguments (47% vs. our 71%; Kuhn et al., 1997) and showed less discourse advancement, achieving, for example, an average level of Counter-C dialogue of 16% (Felton, 2002), compared with our 31%.

The changes observed in discourse strategies are consistent with the earlier cross-sectional study comparing adolescents with a young adult sample (Felton & Kuhn, 2001). Decline in the proportion of discourse devoted to exposition of one's own position reflects an enhanced understanding and implementation of the strategic nature of argumentive discourse. The function of the Clarify discourse strategy is to articulate further one's own argument; it does not address the opponent's argument. The Add strategy bases itself on what the opponent has said, but it merely adds elaboration from the speaker's perspective rather than advancing the argument. An even more direct indication of the development of argumentive discourse skill is increased usage of the strategies that challenge an opponent's claim, from the least complex—Disagree—to the successively more advanced—Counter-A, Counter-C, and the sequence Rebuttal. These skills are central to the goals of argumentive discourse and reflect the shift of attention from exposition of one's own argument to the expanded focus necessary to address the dual goals of argumentive discourse, which include undermining the opponent's argument and advancing one's own argument.

Tempering this positive picture of the change induced in the present work, however, is the fact that the Rebuttal sequence, along with a number of other strategies that are key to skilled argumentive discourse, remain at a relatively low usage level, no greater than 5% for any strategy or sequence. The critical strategy Clarify-?, for example, in which the speaker seeks to understand better the opponent's assertion, increased insignificantly in our sample, from roughly 4% to 5% from first to second

assessment. Other more skilled sequences of strategies, such as Corner and Block, observed by Felton and Kuhn (2001), appeared even less frequently. The emergence of Counter-C and rebuttal in the present study, then, signify at most the beginnings of the realization of the dual goals of skilled argumentive discourse outlined earlier. For use of these strategies to increase, we speculate, requires not only further exercise and development of the skills themselves but also an enhanced metalevel awareness of task goals (Kuhn, 2001a, 2001b): to weaken the opponent's claim or to regain the strength of one's own assertion if it has been undermined by the opponent.

Although this development may take years to accomplish fully, the foundations from which these skills arise have been in preparation for many years. We would not challenge anyone who sought to demonstrate that claims, counterarguments, and rebuttals can be identified in the discourse of young children (Eisenberg & Garvey, 1981). As is the case for many complex cognitive skills, inchoate forms that resemble later, more complex skills can be identified in young children. The important questions, rather, are what exactly is developing and how (Haith & Benson, 1998). In the case of argumentive discourse, commonplace is 3-year-old discourse in the vein of, "My tower is taller." "No, mine is." "No, mine is." What in our view differentiates this inchoate form of argumentive discourse from the argumentive discourse that has been the topic of our work is the epistemological framework in which the discourse is situated (Hofer & Pintrich, 2002; Kuhn, Cheney, & Weinstock, 2001). It is accepted, even by the participants themselves, that toddlers' disputes about towers' tallness are resolvable by resort to empirical verification, with one party necessarily proved correct and the other incorrect, because the opposing positions consist essentially of a claim and its negation. The arguments about a complex social issue such as the one we examined here, in contrast, involve opposing claims whose comparative merit can only be evaluated in a framework of argument and evidence. Relatively late developing is the evaluativist epistemology proposed elsewhere (Kuhn, 2001b; Kuhn et al., 2001) that is needed to support this level of argumentive discourse. Yet it is an essential underpinning, in our view, of the argument schema that Reznitskaya et al. (2001) described. The most salient point to be made in relation to the data presented here is that the developmental goal needs to be more than one of fostering particular discourse strategies. In the words of Anderson et al. (1998), a complementary

and critical goal is to promote "the values and habits of mind to use reasoned discourse as means for choosing among competing ideas" (p. 172). The fact that arguments increased in quantity, as well as quality, in both the present study and the study by Reznitskaya et al. (2001) could be interpreted as an indication that these "habits of mind" were in fact strengthened.

The emergence of a dialogic dimension in participants' individual arguments (Table 8) supports the claim that the two kinds of argument are closely related. The point is an important one for it suggests that a dynamic, dialogic approach is the best way to support the development of skilled argument. If we conceptualize individual arguments supporting a claim as interiorized dialogic argument (Billig, 1987; Kuhn, 1991), the externalization that discourse offers should provide the most effective support for the development of both forms of argument skills.

We are of course not the first to make this claim. Beginning with Vygotsky (1981), numerous authors have emphasized the value of external social collaboration in promoting more advanced forms of individual reasoning. Most notable among researchers to adopt this perspective with respect to argument are Anderson, Reznitskaya, and their colleagues. Reznitskaya et al. (2001) engaged school-aged children in small-group discussions (with 6–10 children per group) of a story and demonstrated that subsequent written arguments in support of a claim showed improvement in quality relative to those of a control group. Reznitskaya et al. did not assess argumentive discourse directly but noted that it would be important to do so. Like current work originating in our laboratory (Felton, in press; Kuhn & DeFuccio, 2002), their work showed, however, that gains are not limited to the topic or content that was the focus of the intervention, a critical step in cognitive intervention research (Sa, West, & Stanovich, 1999; Stanovich, 1999).

Anderson et al. (2001) examined discourse in their 6- to 10-child groups of fourth graders discussing a story. Group size precluded use of the kind of discourse coding scheme employed here and by several other researchers (Keefer et al., 2000; Rips et al., 1999) that tracks the precise relation between one speaker's utterance and another's reply (because in groups it is often not clear whether an utterance is directed at any particular individual or subgroup of individuals). Nonetheless, Anderson et al. (2001) reported increases over time in occurrence of elementary versions of the argument strategies examined here, for example, occurrence of the word

but in discourse as a rudimentary form of counter-argument.

"Collaborative discussion appears to be an effective training ground for the development and internalization of generalized knowledge of argumentation," concluded Reznitskaya et al. (2001, p. 173). If discourse is indeed the social scaffold from which individuals' argumentative reasoning develops, it stands to reason that analysis of its development is of interest not only in its own right but because of the insight it promises into the developing cognitive competence of individuals, a more traditional topic of psychological investigation. In addition, the educational implications of the present work we believe are important, and we explore them more fully elsewhere (Kuhn, in press). In the present context, it is enough to suggest that the developmental research presented here offers a contribution to devising more comprehensive indicators of educational achievement than those represented by the traditional assessment instruments educators continue to rely on so heavily (Bereiter, 2002; Yeh, 2002). Learning and cognitive development indeed come together at this point.

References

- Anderson, R., Chinn, C., Waggoner, M., & Nguyen, K. (1998). Intellectually stimulating story discussions. In J. Osborn & F. Lehr (Eds.), *Literacy for all* (pp. 170–196). New York: Guilford.
- Anderson, R., Chinn, C., Chang, J., Waggoner, M., & Yi, H. (1997). On the logical integrity of children's arguments. *Cognition and Instruction, 15*, 135–167.
- 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.
- Bereiter, C. (2002). *Education and mind in the knowledge age*. Mahwah, NJ: Erlbaum.
- Berkowitz, M., & Gibbs, J. (1983). Measuring the developmental features of moral discussion. *Merrill-Palmer Quarterly, 29*, 399–410.
- Billig, M. (1987). *Arguing and thinking: A rhetorical approach to social psychology*. Cambridge, England: Cambridge University Press.
- Brem, S., & Rips, L. (2000). Evidence and explanation in informal argument. *Cognitive Science, 24*, 573–604.
- Chambliss, M., & Murphy, P. K. (2002). Fourth and fifth graders representing the argument structure in written texts. *Discourse Processes, 34*, 91–115.
- Chinn, C., & Anderson, R. (1998). The structure of discussions that promote reasoning. *Teachers College Record, 100*, 315–368.
- Clark, R., & Delia, J. (1976). The development of functional persuasive skills in childhood and early adolescence. *Child Development, 47*, 1008–1014.
- Eisenberg, A., & Garvey, C. (1981). Children's use of verbal strategies in resolving conflicts. *Discourse Processes, 4*, 149–170.
- Felton, M. (in press). The development of discourse strategies in adolescent argumentation. *Cognitive Development*.
- Felton, M., & Kuhn, D. (2001). The development of argumentative discourse skills. *Discourse Processes, 32*, 135–153.
- Haith, M., & Benson, J. (1998). Infant cognition. In W. Damon (Series Ed.) & D. Kuhn & R. Siegler (Vol. Eds.), *Handbook of child psychology: Vol II. Cognition, perception, and language* (pp. 199–254). New York: Wiley.
- Hidi, S., Berndorff, D., & Ainley, M. (2002). Children's argument writing, interest and self-efficacy: An intervention. *Learning and Instruction, 12*, 429–446.
- Hofer, B., & P. Pintrich (Eds.). (2002). *Epistemology: The psychology of beliefs about knowledge and knowing*. Mahwah, NJ: Erlbaum.
- Keefer, M., Zeitz, C., & Resnick, L. (2000). Judging the quality of peer-led student dialogues. *Cognition and Instruction, 18*, 53–82.
- Klaczynski, P. (2000). Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: A two-process approach to adolescent cognition. *Child Development, 71*, 1347–1366.
- Knudson, R. (1992). Analysis of argumentative writing at two grade levels. *Journal of Educational Research, 85*, 169–179.
- Kuhn, D. (1991). *The skills of argument*. Cambridge, England: Cambridge University Press.
- Kuhn, D. (1995a). Microgenetic study of change: What has it told us? *Psychological Science, 6*, 133–139.
- Kuhn, D. (Ed.). (1995b). Development and learning: Reconceptualizing the intersection [Special issue]. *Human Development, 38*(6).
- Kuhn, D. (2001a). Why development does (and doesn't) occur: Evidence from the domain of inductive reasoning. In R. Siegler & J. McClelland (Eds.), *Mechanisms of cognitive development: Neural and behavioral perspectives* (pp. 221–249). Mahwah, NJ: Erlbaum.
- Kuhn, D. (2001b). How do people know? *Psychological Science, 12*, 1–8.
- Kuhn, D. (in press). *Education for thinking*. Cambridge, MA: Harvard University Press.
- Kuhn, D., Cheney, R., & Weinstock, M. (2001). The development of epistemological understanding. *Cognitive Development, 15*, 309–328.
- Kuhn, D., & DeFuccio, M. (2002). *Teaching thinking among incarcerated youth*. Unpublished manuscript, Teachers College Columbia University.
- Kuhn, D., & Lao, J. (1996). Effects of evidence on attitudes: Is polarization the norm? *Psychological Science, 7*, 115–120.
- Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentative reasoning. *Cognition and Instruction, 15*, 287–315.

- Lao, J., & Kuhn, D. (2002). Cognitive engagement and attitude development. *Cognitive Development, 17*, 1203–1217.
- Mason, L. (1998). Sharing cognition to construct scientific knowledge in school contexts: The role of oral and written discourse. *Instructional Science, 26*, 359–389.
- Mason, L. (2001). Introducing talk and writing for conceptual change: A classroom study. In L. Mason (Ed.), *Instructional practices for conceptual change in science domains. Learning and Instruction, 11*, 305–329.
- Orsolini, M. (1993). "Dwarfs don't shoot": An analysis of children's justifications. *Cognition and Instruction, 11*, 281–297.
- Perkins, D. (1985). Post-primary education has little impact upon informal reasoning. *Journal of Educational Psychology, 77*, 563–571.
- Pontecorvo, C., & Girardet, H. (1993). Arguing and reasoning in understanding historical topics. *Cognition and Instruction, 11*, 365–395.
- 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.
- Rips, L. (2002). Circular reasoning. *Cognitive Science, 26*, 767–795.
- Rips, L., Brem, S., & Bailenson, J. (1999). Reasoning dialogues. *New Directions in Psychological Science, 8*, 172–177.
- Sa, W., West, R., & Stanovich, K. (1999). The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology, 91*, 497–510.
- Siegler, R. (2000). The rebirth of children's learning. *Child Development, 71*, 26–35.
- Siegler, R., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist, 46*, 606–620.
- Siegler, R., & Svetina, M. (2002). A microgenetic/cross-sectional study of matrix completion: Comparing short-term and long-term change. *Child Development, 73*, 793–809.
- Stanovich, K. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.
- Stein, N., & Miller, C. (1993). The development of memory and reasoning skill in argumentative contexts: Evaluating, explaining, and generating evidence. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 285–335). Hillsdale, NJ: Erlbaum.
- van Eemeren, F., Grootendorst, R., & Henkemans, F. (1996). *Fundamentals of argumentation theory: A handbook of historical backgrounds and contemporary developments*. Mahwah, NJ: Erlbaum.
- Voss, J. (Ed.). (2001). Argumentation in psychology [Special issue]. *Discourse Processes, 32*(2 & 3).
- Voss, J., & Means, M. (1991). Learning to reason via instruction in argumentation. *Learning and Instruction, 1*, 337–350.
- Voss, J., & Van Dyke, J. (2001). Argumentation in psychology: Background comments. *Discourse Processes, 32*, 89–111.
- Voss, J., & Wiley, J. (1997). Developing understanding while writing essays in history. *International Journal of Educational Research, 27*, 255–265.
- Vygotsky, L. (1981). The genesis of higher mental functions. In J. Wertsch (Ed.), *The concept of activity in Soviet psychology* (pp. 144–188). Armonk, NY: Sharpe.
- Walton, D. N. (1989). Dialogue theory for critical thinking. *Argumentation, 3*, 169–184.
- 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.
- Yeh, S. (2002). Tests worth teaching to: Constructing state-mandated tests that emphasize critical thinking. *Educational Researcher, 30*, 12–17.
- 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

Summary of Discourse Codes (from Felton & Kuhn, 2001)

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 (rather than 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.
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.	Interpret	A paraphrase of the partner's preceding utterance with or without further elaboration.
Clarify	A clarification of speaker's own argument in response to the partner's preceding utterance.	Meta	An utterance regarding the dialogue itself (rather than its content).
Coopt	An assertion that the partner's immediately preceding utterance serves the speaker's opposing argument.	Null	An unintelligible or off-task utterance.
Counter-A	A disagreement with the partner's preceding utterance, accompanied by an alternate argument.	Refuse	An explicit refusal to respond to the partner's preceding question.
Counter-C	A disagreement with the partner's preceding utterance, accompanied by a critique.	Substantiate	An utterance offered in support of the partner's preceding utterance.
Disagree	A simple disagreement without further argument or elaboration.	Continue	A continuation or elaboration of the speaker's own last utterance that ignores the partner's immediately preceding utterance.
Dismiss	An assertion that the partner's immediately preceding utterance is irrelevant to the speaker's position.	Unconnected	An utterance having no apparent connection to the preceding utterances of either partner or speaker.