

Dialogic Argumentation as a Vehicle for Developing Young Adolescents' Thinking

Deanna Kuhn and Amanda Crowell

Teachers College, Columbia University

Psychological Science
22(4) 545–552
© The Author(s) 2011
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797611402512
http://pss.sagepub.com



Abstract

Argumentative reasoning skills are featured in the new K–12 Common Standards (Common Core State Standards Initiative, 2010), yet with little said about their nature or how to instill them. Distinguishing reasoning skills from writing skills, we report on a multiyear intervention that used electronically conducted dialogues on social issues as the medium to develop argumentative reasoning skills in two cohorts of young adolescents. Intervention groups demonstrated transfer of the dialogic activity to two individual essays on new topics; argument quality for these groups exceeded that of comparison groups who participated in an intervention involving the more face-valid activity of extensive essay writing practice, along with whole-class discussion. The intervention group also demonstrated greater awareness of the relevance of evidence to argument. The dialogic method thus appears to be a viable one for developing cognitive skills that the comparison-group data show do not routinely develop during this age period.

Keywords

argumentation, dialogue, collaborative reasoning, epistemology, adolescence

Received 8/7/10; Revision accepted 11/16/10

Argument has been referred to as the umbrella under which all reasoning lies—“the more general human process of which more specific forms of reasoning are a part” (Oaksford, Chater, & Hahn, 2008, p. 383). Notable among the new K–12 Common Standards (Common Core State Standards Initiative, 2010) is the standard that students become proficient in “logical arguments based on substantive claims, sound reasoning, and relevant evidence.” Not further specified is the nature of this reasoning or how the standard might be achieved. Empirical research may thus provide a foundation for achievement of such a standard.

In this article, we present evidence that argumentative reasoning skills can be identified, assessed, and, most important, developed in facilitative settings. We focus on the middle-school years as an optimal period to undertake this effort, and we follow Graff (2003), and before him the sociocultural tradition of Vygotsky (1978), in taking the everyday social practice of argumentation as a starting point and pathway for development of individual argumentative reasoning. Dialogic argument, Graff suggested, provides the “missing interlocutor” that gives written argument a point. Too often, he claimed, students see the latter only as an exercise in which one strings together a set of reasonable-sounding statements, being careful not to include anything that anyone might challenge. From a developmental perspective, a virtue of a dialogic approach is its foundation in children’s everyday conversations. A familiar activity—everyday talk—has the potential to develop into a more formal, symbolic, and intrapersonal one (Kuhn, 1991).

In the extended intervention we examined, the medium of discourse was electronic. Adolescents argued with opponents using chat software of the sort that most were familiar with and already comfortable using. Beyond the familiarity factor, the electronic medium had the advantage of providing a record of the exchange that remained available throughout and following the discourse. Contributions to face-to-face discourse, in contrast, disappear as soon as they are spoken. In addition to serving as a reference point and framework during the dialogues, these transcripts became the object of various reflective activities participants engaged in.

Although dialogic assessments also showed significant effects, we report here on individual essays as outcome measures because they reflect transfer of skill from the social to the individual plane and also because they are measures of greatest familiarity and concern to educators. We take care to distinguish effects of the intervention from effects of more general school climate and from cognitive development more broadly during this age period (via matched within-school and external comparison groups). We also incorporate the strengths of both an experimental and a repeated measures longitudinal design. The repeated measures design establishes that advances over time in the intervention group were not paralleled by

Corresponding Author:

Deanna Kuhn, Box 119, Teachers College, Columbia University, New York, NY 10027

E-mail: dk100@columbia.edu

comparable changes in the comparison group. Precise longitudinal assessment, however, demands use of the same measure across assessments, and this raises the possibility of practice effects that may blur the longitudinal picture. We therefore administered two similar assessments—one across occasions, to track change over time within groups, and the other only at the end of the intervention as a basis for comparing groups on an assessment not previously encountered.

The distinction between what participants were engaged in during the intervention and their task on the assessments warrants emphasis. Intervention participants discussed only four topics each year, each twice weekly for approximately a 7-week period; they engaged deeply with the topics, debating them with same-side and opposing-side classmates in various configurations and, in the 2nd year, accessing and bringing to bear relevant evidence. By the end of a topic cycle, which culminated in a final whole-class debate, debriefing, and individual essay, they had addressed the topic intensively.

The assessment essays, in contrast, were on a topic participants were assigned only once (posttest-only assessment) and another topic they were assigned four times (at yearlong intervals). Except by chance, participants did not contemplate these topics elsewhere. Hence, gains in the comparison group could be attributed to greater knowledge or reasoning they could bring to bear on a topic; any additional gains in the experimental group could be attributed to the intervention. Such effects, however, could be only indirect, as the intervention did not involve the assessment topics.

We gave considerable thought to the competencies we wished the essays to assess, drawing on the thinking of Toulmin (2003) and Walton (1989), as well as the sociocultural perspective noted earlier. Although the need to improve students' writing skills has long been an educational concern, the distinction between thinking skills and writing skills has not been clearly drawn. Our focus here is the development of thinking, rather than writing, skills. At its most minimal level, thinking well about a complex issue requires identifying and weighing positive and negative attributes of contrasting positions on the issue, drawing on relevant evidence to inform the judgments involved.

Most students' persuasive essays fall well short of this standard (National Assessment of Educational Progress, 2008); most often they are limited to presenting positive attributes of the favored position. A distant second in frequency is exposition of negative attributes of an opposing position. If some combination of both appears, an argument can be classified as reflecting what we call here a *dual* perspective, so termed because the arguer must shift at least once from positive to negative attributes and from the perspective of the favored position to that of the opposing position. A dual perspective is also significant in that it reflects *counterfactual* reasoning; that is, it requires assuming a stance contrary to one's own and reasoning about its implications.

Elementary arguments may also include exposition of negative attributes of the favored position and exposition of positive attributes of the opposing position. Including either of these requires the arguer to exhibit what we call an *integrative*

perspective. In contrast to the dual perspective, in which all arguments lead to the same conclusion, the integrative perspective gives voice to a set of arguments that lead in opposing directions and hence require an integrative weighing in order for a conclusion to be reached. In the Results section, we elaborate on how these concepts were applied in coding the essays that constituted our data set.

Method

Participants

Participants were entering sixth graders (all 11 or 12 years old) attending an academically challenging urban public middle school in an ethnically diverse, low- to middle-income neighborhood in the Northeast United States. Eighty percent were Hispanic or African American, and 60% qualified for free or reduced-price lunch.

The school assigned entering students to one of three 30- to 32-student classes equated for gender, ethnicity, and standardized-test scores. Two classes were randomly selected as our experimental group, and the third served as the comparison group.

The final longitudinal sample contained 48 students in the experimental group (27 female, 21 male) and 23 (11 female, 12 male) in the comparison group. All participants completed at least three of the four assessments (initial and end of Years 1, 2, and 3), except for 1 (in the comparison group) who missed both interim assessments. Three participants (1 experimental, 2 comparison) missed the final assessment, and 7 missed one interim assessment (5 experimental, 2 comparison). The others completed all four assessments.

The second (posttest-only) essay assessment in Year 3 allowed inclusion of students who completed the final assessment but were not included in the longitudinal sample as a result of having missed the initial assessment, either because of absence or because they did not enter the school until between the middle of Year 1 and the beginning of Year 3 (almost all students in this category entered at the beginning of Year 2). For this analysis, the sample size increased to 60 students in the experimental group (31 female, 29 male) and 28 in the comparison group (13 female, 15 male).¹

As an additional comparison, we secured at the end of Year 3 an external comparison group of 50 eighth graders (roughly half female and half male) from a public school in the same city, to whom we administered the posttest-only assessment. The selected school was closely matched to the main-sample school on standardized scores, percentage of students eligible for free or reduced-price lunch, and percentage of students identified as African American or Hispanic.

Assessments

At the beginning of Year 1 (Y0) and end of Years 1, 2, and 3 (Y1, Y2, and Y3, respectively), the longitudinal sample was asked to respond in writing to this prompt:

The new Columbia Town School has to decide how to pay its teachers. Some think every teacher should get the same pay. Others think that teachers should be paid according to how much experience they have, with teachers getting more pay for each year of teaching experience they have. Which do you think is the better plan and why?

At the end of Y3, all participants also responded in writing to this prompt (posttest-only assessment):

Sometimes those with an incurable illness want to end their own lives. Should doctors and family members be allowed to assist them? Why or why not?

In addition to assessing participants' argumentative reasoning, we sought to assess their understanding of the role of evidence in argument. Accordingly, at Y3, following each essay, participants were asked to respond to an additional prompt:

Are there any questions you would want to have answers to that would help you make your argument? List them below.

The Y3 assessments were group-administered several days apart, with the teacher-pay assessment coming first. All participants reported that they had finished in the allotted time of 20 min for the essay and 10 min for listing their questions.

Intervention

The two intervention classes met as intact groups for a twice-weekly 50-min class, identified as a philosophy class. Each school year was divided into four quarters of about 13 class sessions each. A unique topic was introduced each quarter as the basis for that quarter's work. A 1st-year topic, for example, was whether parents should be allowed to home-school a child, a 2nd-year topic was China's one-child policy, and a 3rd-year topic concerned whether adult court, juvenile court, or teen (peer judge) court was the best means for adjudicating juvenile crimes. (All 3rd-year topics were three-sided.) Participants chose their sides on a topic. Topics had been pilot-tested to achieve approximately equal numbers of students who favored the different sides.

The topic cycle began with small-group team work among students on the same side (*pregame*) and proceeded to electronic dialogues between pairs of students on opposing sides (*game*). Next, small-group preparation preceded a whole-class verbal debate that served as the capstone experience of the sequence (*endgame*). A debriefing session concluded with a final individual essay assignment.²

Pregame (Sessions 1–3). During the pregame phase, participants met in same-side groups of 7 to 8, each with an adult coach who acted only to facilitate group process. The first

session ("Our Reasons") was devoted to generating reasons why the position the group favored was the better one and assembling a set of "reason cards" that represented their supporting reasons. The second session ("Evaluating Reasons") focused on evaluation and ranking of reason cards with respect to their strength as support for the position.

Beginning with the final topic of Year 1 and continuing during Years 2 and 3, participants were offered relevant evidence that they were told might help support their group's position. Although participants showed little concern with evidence during Year 1, it became an increasingly important focus during the following years. Beginning with Session 2 for Topic 4 in Year 1, evidence was introduced by means of a set of 8 to 10 envelopes. A question was typed on the outside of each envelope, and a two- to three-sentence answer appeared inside. For example, for the Year 2 animal-rights topic, two questions were, "How humanely are animals treated in laboratories?" and "Has animal testing led to cures for any human diseases?" Beginning in Year 2, participants were also invited to submit their own questions, and by the middle of that year, the number of questions they submitted outstripped the number created by us. By the end of Year 2, participants generated all questions themselves, reaching a high of 50 distinct questions for a topic in Year 3. Coaches supplied two- to three-sentence answers to participant-generated questions by the following session, although in Year 3 some participants volunteered to assist in Internet research to obtain answers. All question-and-answer materials remained accessible throughout the topic cycle.

The third session ("Others' Reasons") focused on anticipating what the other side's reasons might be, considering how those reasons might be countered, anticipating how the other team would counter one's own side's reasons, and conceiving of ways to rebut these counters ("comebacks"). These activities were supported by construction of "super cards," each consisting of a reason card stapled together with various additional cards (evidence cards, counter cards, and rebuttal cards). "Other-side cards," containing counterarguments to the other side's arguments, were also constructed.

Game (Sessions 4–9). Each participant was paired with a specific same-side peer throughout the game phase. Together, the pair argued against a sequence of six opposing-side pairs, one each session. Dialogues took place electronically via Google Chat. Participants were reminded to collaborate with their partners in deciding what to say to the opposing pair. Each dialogue lasted approximately 25 min.

During each session, while waiting for the opposing pair to respond, partners completed a reflection sheet, referring to the ongoing dialogue transcript that appeared on the screen before them. The reflection sheets came in two forms (alternated across sessions). One asked the pair to identify and reflect on one of their own arguments, and the other asked them to reflect on one of the opponents' arguments. Each asked for identification of counters and rebuttals and whether these could be improved. In Year 3, reflection sheets were replaced by an assignment asking

students to reflect on some of their dialogue transcripts and evaluate them in writing on a move-by-move basis.

Endgame (Sessions 10–12). Participants returned to same-side small groups and engaged in two sessions of preparation for the final “showdown” debate. One session focused on reviewing the arguments the other side used in the dialogues and the counterarguments that could be used against them in the showdown. In the other session, students focused on their own side’s arguments, expected counterarguments, and rebuttals.

For the showdown session, participants remained in their small groups; the first half of the session focused on two of the small groups (one from each side), and the second half on the other two. As decided on by the group, one member at a time went to a “hot seat” and verbally debated someone from the opposing side. Time was called after 3 min, and a new teammate came to the hot seat. Students in the inactive small groups were able to observe and forward written suggestions to the active groups. Whenever they wished, members of either active team were able to call a 1-min “huddle” to confer.

Debriefing and final assignment (Session 13). In a debriefing session, participants were guided through an argument map—a diagram of the debate with points awarded for effective argumentative moves (typically counterarguments and rebuttals) and points subtracted for ineffective moves, such as

unwarranted assumptions and unconnected responses. Points were summed and a winner declared. Finally, participants were assigned to write individual essays justifying their positions on the topic; these essays were due at the next session, at which time a new topic cycle began.

Comparison group

The comparison group similarly met as an intact group for a twice-weekly 50-min philosophy class, taught by a teacher from the school. This class covered a larger number of topics than the experimental classes, but all topics involved philosophical or social issues, including some of the issues addressed by the experimental classes. Students engaged in teacher-led, whole-class discussion of the issues, along with some additional activities such as dramatizations, and were assigned essays at least once every 2 weeks. Hence, they obtained more practice in writing expository essays on such topics than did the experimental group (14 per year vs. 4).

Results

Assessment essays were divided into idea units, and each idea unit was classified into one of four categories (no argument, own-side argument, dual-perspective argument, or integrative-perspective argument). Table 1 provides verbatim examples.

Table 1. Coding Scheme for the Teacher-Pay Essay

Argument type	Examples	
	Equal pay is favored position	Experience-based pay is favored position
No argument	All the teachers should get the same pay because it shouldn't depend on how experienced you are as a teacher that determines how much you get paid.	More experienced teachers should get paid more because if a new teacher just came into the school and another teacher has been working there for years the more experienced teacher should get more money.
Own-side only (includes only positives of the favored position)	Teachers should get paid the same amount because they are all going to teach a subject that is going to help the children's education in some way. I think all teachers should get the same pay because think how hard ALL teachers work.	Teachers with more experience should get paid more because they are the ones that worked hard to get in the position they're in. Experienced teachers should get more pay because of the skills they have, because they have more to offer the students.
Dual perspective (includes negatives of the opposing position)	If teachers were paid according to experience, this would create conflict for the teachers because there would be a very large disagreement on how much each teacher is getting paid. Unequal pay wouldn't be good because experienced teachers have already been paid for their previous years of teaching; it would be like paying them twice. The new teachers might not even want to teach at a school that gives them so little pay; then how will you get new teachers?	If new teachers got the same pay, experienced teachers would get fed up and quit. If experienced teachers got the same pay as new teachers, they would feel like it was unfair and not want to help out the new teachers.
Integrative perspective (includes negatives of the favored position or positives of the opposing position)	Experienced-based pay may seem fair to those who have taught for a long time [positive of the opposing position]. But not for the new teachers who do just as much as everyone else [negative of the opposing position].	Although it does seem unfair the school is basing your salary on age [negative of the favored position], it's a clever way to keep good teachers for a longer time [positive of the favored position]. This cannot be viewed as unfair to new teachers [negative of the favored position] because with more experience they will receive the pay they deserve [positive of the favored position].

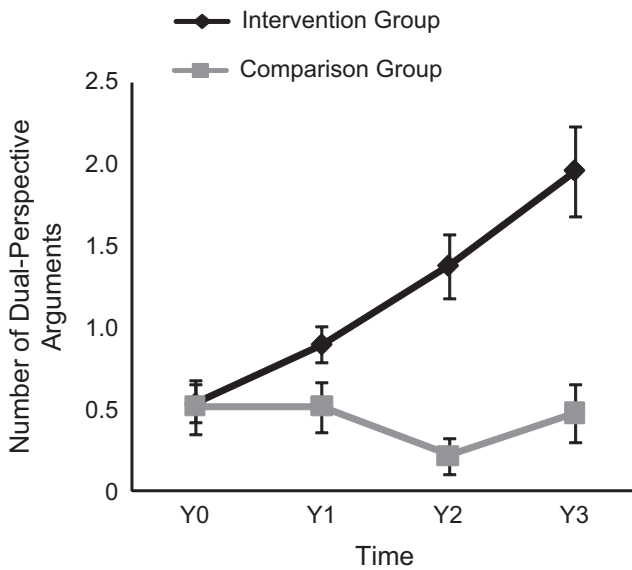


Fig. 1. Mean number of dual-perspective arguments in the teacher-pay essay as a function of time and condition. Y0 is the beginning of the study, and Y1, Y2, and Y3 are the ends of the 1st, 2nd, and 3rd years of the study, respectively. Error bars represent standard errors of the means.

Classification was made blind to condition and time; a randomly chosen third of the essays were coded by a second coder. For the teacher-pay essay, percentage agreement was 88%, Cohen’s $\kappa = .76$; for the euthanasia essay, it was 93%, Cohen’s $\kappa = .91$.

Teacher-pay essays

Figure 1 presents the mean number of dual-perspective arguments in the teacher-pay essays as a function of time and condition. There was a significant interaction between condition and time, $F(3, 67) = 6.11, p < .001, \eta_p^2 = .216$. Simple-effects tests showed that the experimental group included more dual-perspective arguments than the comparison group at Y2, $t(69) = 3.60, p = .001$, and Y3, $t(69) = 3.89, p < .001$. Table 2 shows the percentages of participants who made any dual-perspective or integrative-perspective arguments. No integrative arguments appeared until the Y3 assessment.

Essays became longer over time in both conditions. Thus, gains in the experimental condition cannot be attributed simply

to increasing verbal productivity. (In the experimental condition, the mean number of words rose from 100 at Y0 to a maximum of 142 at Y2; in the comparison condition, the mean rose from 105 words at Y0 to 151 at Y2.) We also analyzed the total number of arguments, which showed effects of time, $F(3, 67) = 26.00, p < .001, \eta_p^2 = .234$, and condition, $F(1, 69) = 7.37, p = .008, \eta_p^2 = .097$, with no interaction. The main effect of condition was accounted for by a significant condition difference at Y3 only; at this assessment, experimental-group participants offered more arguments than comparison-group participants (3.96 vs. 2.61), $t(69) = 2.81, p < .006$. However, even when we converted the Y3 data in Figure 1 to proportions to adjust for overall differences in production, the condition difference remained robust (.49 vs. .24).³

Euthanasia essays

Groups did not differ significantly on number of words or arguments in the euthanasia essay; however, the experimental group generated more dual-perspective arguments than the internal comparison group did ($M = 1.17$ vs. 0.36), $t(86) = 3.85, p < .001$. Table 3 presents the percentage of students in the experimental and comparison groups who made dual-perspective and integrative-perspective arguments. Also presented are parallel percentages for the external comparison group.

Questions

In addition to counting the number of questions posed for each topic in response to the question probes at Y3, we coded these questions as either case based or general. Case-based questions were those whose answers appeared to have implications for the resolution of the specific case only, whereas general questions were those that had the potential to affect judgments about the issue in general (see Table 4 for examples). Table 5 shows the mean number of questions posed by each group (including the external comparison group). For both topics, the experimental group posed significantly more questions than the internal comparison group did—teacher-pay topic: $t(69) = 7.63, p < .001$; euthanasia topic: $t(86) = 3.32, p = .001$. Moreover, case-based questions were more prevalent in the comparison groups than in the experimental group.

Table 2. Percentages of Participants Who Made Dual-Perspective and Integrative-Perspective Arguments in the Teacher-Pay Essay

Argument type	Initial		End of Year 1		End of Year 2		End of Year 3	
	E	C	E	C	E	C	E	C
Dual perspective	35	35	67	38	79	19	79	29
Integrative perspective	0	0	0	0	0	0	30	0

Note: E = experimental group; C = comparison group.

Table 3. Percentages of Participants Who Made Dual-Perspective and Integrative Arguments in the Euthanasia Essay at the End of Year 3 and Examples of Those Arguments

Type of argument	Examples	Percentage of students		
		Experimental group	Comparison group	External comparison group
Dual perspective	Pro-euthanasia: Why let the person suffer and let the family suffer financially? Anti-euthanasia: If everybody did euthanasia, no progress in medication could be made.	73%	29%	24%
Integrative perspective	Pro-euthanasia: Even though it might not be easy to stand by and watch [negative of the favored position], a family member must support whatever the ill person wants [positive of the favored position]. Anti-euthanasia: Sure it will take away the pain right away [positive of the opposing position], but that wouldn't be the only thing gone [negative of the opposing position].	30%	18%	8%

Table 4. General and Case-Based Questions That Participants Posed About Euthanasia in Response to the Question Probe

General questions	Case-based questions
How many people in a hospital (average) have an incurable disease?	Who is the person? (Adult, child, etc.)
What incurable diseases are there and what are the chances of surviving them?	What is their disease?
What are the most painful diseases?	How old is the patient?
What are the treatment options for incurable diseases (e.g. cancer)?	In what conditions is this person in?
What can doctors do to ease pain for someone?	How long do they have to live?
How many people that have incurable illness go to a therapist?	Is there any way to cure them?
Does depression affect suicide?	Are they in pain?
What percent of people who have an incurable disease want to die?	Does the family care what their decision is?
How many people have asked doctors to let them die?	Are they on life support?
How many people are assisted with killing themselves?	Do they have a family—for example if the women had kids?
How would a life-ending procedure go? For example, would it be a shot?	What do they do for a living—for example if they have an important role in society?
What do the majority of doctors believe on this issue?	
Do doctors get overwhelmed when asked to help someone commit suicide?	
Can doctors oppose the wishes of the patient and its family members?	
Does the family of the patient have any say in this?	
How do families react to the death?	
How much is it (money) to keep the person on life support vs. ending their life?	
What percentage of doctors/hospitals do this?	
What states allow doctors to kill patients?	
How many countries allow doctors to help kill patients?	
Have doctors been sued because of this?	
What are the number of doctors that get put in jail due to assisted suicides?	
Is it illegal to kill yourself?	
Is it considered murder when you kill someone who wanted you to kill them?	
Where is ending someone's life legal?	

Note: These questions and very similar ones are all the questions that participants asked. They were not answered (in contrast to questions participants posed during the intervention).

Replication Study

A replication sample was drawn from the next incoming class. This replication study followed the design and method for the

first 2 years of the main study except that the sample size was reduced, and assessments were available for the comparison group only at Y2.⁴ Also, the question probe was administered to this sample at Y2.

Table 5. Mean Number of Questions Posed in Response to the Question Probe

Group	Teacher-pay topic	Euthanasia topic
Experimental	4.18 (10% case based)	3.26 (0% case based)
Comparison	0.33 (100% case based)	1.64 (74% case based)
External comparison	—	0.92 (30% case based)

The results for the teacher-pay essays were remarkably similar to the results obtained in the main study (see Table 6), with the notable exception that 1 participant at Y1 and 5 (14%) at Y2 made an integrative argument, whereas such arguments did not appear until Y3 in the main study. At Y2, the total number of arguments did not differ between the groups, but the experimental group generated more dual-perspective arguments than the comparison group ($M = 1.05$ vs. 0.43), $t(56) = 2.81$, $p = .007$. Participants in the experimental group asked an average of 1.48 questions, and participants in the comparison group asked an average of 0.90 questions; the difference was not significant ($p = .148$).

Discussion

There is much talk in current education circles about the importance of “21st-century skills.” But these will be significantly incorporated into educational policy and practice only to the extent that they can be explicitly defined and measured. The present work shows that argumentative reasoning skills not only can be identified and assessed with precision, but also can be developed. In so doing, it also shows how empirical research can contribute to articulating educational goals.

Given its extended time investment and stand-alone (non-curriculum-embedded) nature, the method we employed can be regarded as justified only if the value of the skills in question is recognized. The skills we examined are not especially high-level cognitive skills. Yet they arguably are fundamental to the kinds of higher-order thinking that are of increasing importance in the contemporary world. Counterfactual reasoning, dual-perspective reasoning, and integration of opposing

arguments are essential building blocks of sophisticated, nuanced real-world argumentative reasoning. Our comparison-group data offer little if any indication that the skills we identified develop routinely during the age range examined, in the absence of educational interventions that target them. No improvement over time was apparent in the comparison group. The proportion of participants making dual-perspective arguments remained steady at about one third. The comparison group in the replication sample did slightly better than the comparison group in the main sample at Y2, but the lack of comparison with earlier points makes this difference difficult to interpret with any certainty, and the replication data overall are consistent with the main data.

The method of developing argumentative skills that was used in this study has a number of positive attributes. Dialogic argumentation skill has not been a major concern of educators (Michaels, O’Connor, & Resnick, 2008), and for this reason, we report here on outcomes with respect to individual expository writing of the sort that has received much more attention in educational circles. Yet dialogic argumentation skills are arguably of critical importance in their own right in contemporary life. Moreover, they may be a key to the development of the individual expository skills that educators have given more attention to and that continue to pose a significant educational challenge.

The electronic mode of discourse we employed has the merit not only of initial familiarity for most middle-school students, but also of providing an immediately available written record of the dialogue that enables and promotes reflection. The collaborative aspect of the activity further promoted this reflection; current work (Shaenfield, 2011) has shown that outcomes are inferior when participants engage in dialogues alone, without a same-side partner. Furthermore, the sequence within each topic provided a goal structure that motivated engagement. Finally—no small matter—engagement in deliberating significant social issues enhances appreciation of their complexity. We were surprised, for example, that many middle-school students initially accepted sale of human organs as unproblematic; it was even more startling that many of our participants expressed the view that medical research should be conducted not on animals, who are innocent, but on prisoners, because they are guilty.

Table 6. Percentages of Participants Who Made Dual-Perspective and Integrative-Perspective Arguments in the Replication Study

Argument type	Experimental group			Comparison group, end of Year 2
	Initial	End of Year 1	End of Year 2	
Dual perspective	30	54	73	38
Integrative perspective	0	3	14	0

Note: The sample size was 37 for the experimental group and 21 for the comparison group. Data are missing for 2 experimental participants at the initial assessment and 2 experimental participants at the end of Year 1.

An additional component of the intervention that warrants highlighting is its epistemological aspect. The data on participant-posed questions, we believe, indicate that participants acquired not simply a question-asking routine, or habit, but rather an awareness that evidence is relevant to argument—and in particular, arguments about social issues; a prevailing “science as accumulated fact” epistemological stance may make this awareness easier to acquire in the case of science topics (Kuhn, 2010). Like the argument skill manifested in the essays, which did not show significant improvement until Year 2, this evidence seeking developed only slowly (the effect was significant only at Y3), but we expect both the competence and the disposition involved to have far-reaching effects. It is not enough to teach students *how* to argue well; they must develop the intellectual values and dispositions that will support their doing so (Kuhn, Wang, & Li, 2011). This development entails epistemological understanding of what argument indeed consists of, as well as the power it confers.

This complex, multicomponent intervention, of course, requires further experimental dissection so as to isolate its effective components, and this work remains to be done. In the absence of compelling evidence of success of more “quick fix” approaches, the effort appears worthwhile. Approaches that arguably have greater face validity, such as the extensive practice in essay writing engaged in by our comparison groups, appear to be not as effective as our less direct approach. Our goal is to understand more about how the intervention we examined achieves its effects.

Acknowledgments

We thank Jose Maldonado, principal of Columbia Secondary School, whose vision and conviction regarding the worth of this undertaking made it possible. We thank Shawn Yanklowitz for assistance with reliability coding of the essays.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This work was supported by an award from the Institute of Education Sciences.

Notes

1. This increase was roughly proportional across groups. Nonetheless, to be sure, we also analyzed data without the additional participants and obtained virtually identical results. Hence, we include them in the data reported here.

2. The Year 1 protocol deviated slightly from the description given in that one class spent more time on pregame activities and the other spent more time on game activities, so that we could determine whether this difference in time allocation affected performance; we found that it did not. The Year 3 protocol deviated to accommodate the three-sided structure of Year 3 topics.

3. We therefore chose not to convert the data in Figure 1 to proportions, so as to convey a sense of the absolute levels of performance.

4. The size of the replication sample was compromised by one of the nightmares of researchers doing classroom research. At the beginning of the last quarter of Year 2, the school administration decided to change a number of students' class assignments for behavioral-management reasons. As a result, some students in our comparison group moved to the experimental group and vice versa. We excluded from the analysis all students who changed classes.

References

- Common Core State Standards Initiative. (2010). *Common core state standards for English language arts & literacy in history/social studies, science, and technical subjects*. Retrieved from http://www.corestandards.org/assets/CCSSI_ELA%20Standards.pdf
- Graff, G. (2003). *Clueless in academe: How schooling obscures the life of the mind*. New Haven, CT: Yale University Press.
- Kuhn, D. (1991). *The skills of argument*. Cambridge, England: Cambridge University Press.
- Kuhn, D. (2010). Teaching and learning science as argument. *Science Education, 94*, 810–824.
- Kuhn, D., Wang, Y., & Li, H. (2011). Why argue? Developing understanding of the purposes and value of argumentative discourse. *Discourse Processes, 48*, 26–49.
- Michaels, S., O'Connor, C., & Resnick, L. (2008). Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. *Studies in Philosophy and Education, 27*, 283–297.
- National Assessment of Educational Progress. (2008). *The nation's report card: Writing 2007*. Washington, DC: U.S. Department of Education.
- Oaksford, M., Chater, N., & Hahn, U. (2008). Human reasoning and argumentation: The probabilistic approach. In J. Adler & L. Rips (Eds.), *Reasoning: Studies of human inference and its foundations* (pp. 383–413). New York, NY: Cambridge University Press.
- Shanfield, D. (2011). *Arguing with peers: Examining two kinds of discourse and their cognitive benefits*. Manuscript submitted for publication.
- Toulmin, S. (2003). *The uses of argument* (2nd ed.). Cambridge, England: Cambridge University Press.
- Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Walton, D. (1989). Dialogue theory for critical thinking. *Argumentation, 3*, 169–184.