

**Abstract Title Page**

*Not included in page count.*

**Title:** Argumentation as a Path to the Thinking Development of Young Adolescents

**Author(s):** Deanna Kuhn, Teachers College, Columbia University  
Amanda Crowell, Teachers College, Columbia University

## **Abstract Body**

*Limit 5 pages single spaced.*

### **Background/context:**

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 (2010) rapidly being adopted by states is reference to argumentative reasoning. One standard is 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 nor how the standard might be achieved. Empirical research on reasoning and its development thus has a contribution to make.

We present evidence here that argumentative reasoning skills can be assessed and developed in facilitative settings. We focus on middle school as an optimal period to undertake this effort, and we follow Graff (2003), and before him the sociocultural tradition of Vygotsky (1973) and others, in taking the everyday social practice of argumentation as a starting point and pathway for development of individual argumentative skill (Kuhn, 1991). Dialogic argument, Graff suggests, provides the “missing interlocutor” that gives written argument a point. Too often, he claims, 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.

In the extended intervention reported on here, the medium of discourse is electronic. Students argue using software that most are familiar with and already comfortable using. Beyond the familiarity factor, use of an electronic medium has the significant advantage of providing a transcript of the exchange that remains 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 dialogs, these transcripts become the object of various reflective activities participants engage in.

### **Purpose / objective / research question / focus of study:**

We focus here on individual essays because they are arguably the most powerful outcome measures – of transfer of skill from the social to the individual plane – and also because they are the measures of most familiarity and direct interest to educators. Although persuasive writing has long been a curricula concern, the distinction between thinking skills and writing skills is often blurred. Our concern is the thinking skills that underlie writing. At its most minimal level, thinking well about a complex issue can be regarded as requiring the identification and weighing of positive and negative attributes of contrasting positions on the issue, drawing on relevant evidence to inform the judgments involved. Most students’ persuasive essays, assessments have shown, fall well short of this standard (NAEP, 2008), with most confined to arguments citing 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 a *dual* perspective, since 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. This characteristic is also significant in reflecting *counterfactual* reasoning, since it requires assuming a stance contrary to one’s own and reasoning about its implications.

Two further possibilities exist as attributes of elementary arguments -- exposition of negative attributes of the favored position and/or positive attributes of the opposing position. Either requires the arguer to exhibit what we call an *integrative* perspective. In contrast to the dual perspective, where all arguments lead to the same conclusion, the set of arguments voiced in this case lead in disparate directions and hence require an integrative weighing in order for a conclusion to be reached. These criteria were applied in a coding system for the essays that constitute our data set.

**Setting:**

The intervention took place as a twice-weekly class at an academically challenging urban public middle school in an ethnically diverse low- to- middle-income neighborhood.

**Population / Participants / Subjects:**

Participants were entering 6<sup>th</sup> graders (all 11 or 12 years old) at the beginning of the three-year intervention. Eighty percent were Hispanic or African-American, and 60% qualified for free or reduced-price lunch. The final sample contained 48 in the experimental group (27 female) and 23 (11 female) in a comparison group. As an additional comparison, we secured at the end of year 3 an external comparison group of 50 8<sup>th</sup>-graders (roughly half female) from a public school in the same city, to whom we administered the final essay assessment. The school was selected as closely matched to the main-sample school on test scores, percentage eligible for free or reduced-price lunch, and percentage identified as African-American or Hispanic.

**Intervention / Program / Practice:**

During the 3-year period (Y1-3), the experimental group met as two intact classes for a twice-weekly 50-min class. Each 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 (e.g., in Y1 whether parents should be allowed to home school a child, in Y2 China's one-child policy). Each topic cycle repeated the same sequence of activities.

*Pre-game (Sessions 1-3).* Students met in same-side groups of 7-8 (with students choosing a preferred side), each with an adult coach who acted only to facilitate group process. The first ("Our Reasons") session was devoted to generating reasons why the position the group favored is the better one and assembling a set of "reason cards" that represented their supporting reasons. The second ("Evaluating Reasons") focused on evaluation and ranking of reason cards with respect to their strength as support for the group's position. The third ("Others' Reasons") session focused on a) anticipating what the other side's reasons might be; b) how they might be countered; b) anticipating how the other team will counter our reasons; and d) conceiving of ways that these counters can be rebutted ("comebacks").

Beginning with the final Y1 topic and continuing during Y2-3, relevant evidence was introduced as possibly helping to support the group's position. Students initially showed little concern with evidence, but it became an increasingly important focus during Y2-3. By the end of Y2, they generated all questions themselves, though coaches continued to supply answers.

*Game (Sessions 4-9).* Students were paired with the same same-side peer throughout this phase. Together, the pair argued electronically via *Google-chat* against a sequence of six opposing-side pairs, one per session. Pairs were reminded to collaborate with their partners in deciding on their input to the opposing pair. Each dialog lasted approximately 25 minutes. While waiting for the opposing pair to make their response, the pair was asked to complete one

reflection sheet per session, referring to the ongoing dialog transcript that appeared on the screen before them. These were of one of two forms (alternated across sessions) – one asking the pair to identify and reflect on one of their own arguments and the other on an opponents’ argument.

*End-game (Sessions 10-13).* Students returned to their same-side small groups and engaged in two sessions of preparation for a final “Showdown” whole-class debate. One session focused on reviewing the other-side arguments encountered in the dialogs and the counterarguments against them to use in the Showdown. The other session focused on reviewing own arguments, expected counterarguments, and rebuttals.

For the showdown, participants remained in their small groups, with the first half of the session involving two small groups (one from each side), and the second half the other two. One member at a time was chosen by the small group to come to a “hot seat” and verbally debate a counterpart from the opposing side for three minutes. Members of either active team were able to call a one-minute “huddle” to confer with their teammates.

In a final whole-class debrief session, students were guided through an argument map -- a transcription of the debate with points awarded for effective argumentative moves (notably counterarguments and rebuttals) and points subtracted for ineffective moves, such as unwarranted assumptions and unconnected statements. Points were summed and a winner declared. Finally, students were assigned to write individual final essays justifying their positions on the topic. At the next session they turned these in and a new topic cycle began.

*Comparison group.* The comparison group similarly met as an intact class twice weekly. The class was taught by a teacher from the school. It covered a larger number of topics than the experimental classes, but all topics involved a social issue, some the same ones addressed by the experimental classes. Students engaged in teacher-led, whole-class discussion of the issue, with some additional activities such as dramatizations, and were assigned individual essays on a topic at least once every two weeks. Hence they obtained more practice in writing expository essays on such topics than did the experimental group (14 per year vs. 4).

### **Research Design:**

An extended intervention study demands careful attention to research design. During the 3-year period of the study, young adolescents’ cognitive skills and knowledge are expanding in multiple ways. We therefore distinguish effects of the intervention from effects of more general school climate (via a carefully matched within-school comparison group), and from more general cognitive development occurring during this period (via a matched group from another school).

We also incorporate the strengths of both an experimental and a repeated-measures longitudinal design. The repeated-measures design establishes that the advances observed over time in the intervention group are not paralleled by comparable changes in the comparison group. Precise longitudinal assessment however, demands use of the same measure across assessments, which raises the possibility of practice effects that risk blurring the longitudinal picture. We therefore administered two parallel assessments – one across occasions, to track change over time in both 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 students did during the intervention and their task on the assessment warrants emphasis. Intervention participants engaged only four topics each year, engaging deeply with it, debating it with same-side and opposing-side classmates in various configurations, and, beginning in Y2, accessing and bringing to bear relevant evidence. By the

end of the cycle, culminating in a final whole-class debate, debrief, and individual essay, they were very familiar with the topic.

The essays on which our analysis is based, in contrast, were on a topic students encountered either once only, for one topic, or four times (initially and at the end of each year) for the other. They did not, except by chance, contemplate these topics outside of these specific in-class essay assignments. Hence, gains in the comparison group could be attributed to greater knowledge and/or reasoning they were able to bring to bear on the topic; any additional gains in the experimental group could be attributed to the intervention. Such effects, however, could only be of an indirect type, since the intervention did not involve engagement with these topics.

### **Data Collection and Analysis:**

At the beginning of year 1 (Y0), and ends of Y1, Y2, Y3, students responded in writing to the following 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 students (including the external sample) also responded to this prompt:

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?

Assessments were group-administered several days apart, the teacher pay (TP) assessment first. All students reported they had finished in the allotted time of 20 min.

At Y3, for each essay another 10 min was allowed 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.

Essays were divided into idea units, and each idea unit was classified into one of four categories (no argument, own-side, dual perspective, integrative perspective). Verbatim examples appear in table 1. Classification was made blind to condition and time; a randomly chosen third were coded by a second coder. For TP, percentage agreement was 88%, Cohen's Kappa = .76; for Euthanasia it was 93%, Cohen's Kappa = .91.

### **Findings / Results:**

*Teacher Pay.* Figure 1 presents mean number of dual-perspective arguments by time and condition. There was a significant interaction between condition and time,  $F(3,67)=6.11$ ,  $p<.001$ . Simple effect tests showed the experimental group exceeded the control group at Y2,  $t(69) = 3.60$ ,  $p = .001$ , and Y3,  $t(69) = 3.89$ ,  $p<.001$ . Table 2 shows percentages of participants who made any dual-perspective arguments. No integrative arguments appeared until Y3.

Essays became longer over time but in both conditions (from a mean of 100 words at Y0 to 142 words at Y2, their longest, in the experimental condition and from 105 at Y0 to 151 at Y2 in the comparison condition), arguing against a claim that advances in the experimental condition are attributable only to increasing verbal productivity. As a further test of this possibility we analyzed total number of arguments by time and condition, which yielded effects of both with no interaction, for condition  $F(1, 69)=7.37$ ,  $sig=.008$ ,  $partial \eta^2 = .097$ , for time,  $F(3,67)=26.00$ ,  $sig<.001$ ,  $partial \eta^2 = .234$ . In individual comparisons, the condition main effect was accounted for by a significant condition difference only at Y3, with experimental participants offering more arguments than comparison participants at this assessment, (3.96 vs. 2.61;  $t(69)=2.81$ ,  $p<.006$ ).

However, even if the Y3 data in figure 1 are converted to proportions to adjust for overall production differences, the condition difference remains robust (.49 vs. .24). (We therefore chose not to convert all data to percentages, so as to convey a sense of absolute levels of performance.)

*Euthanasia.* Groups did not differ significantly on number of words or arguments; however, the experimental group generated more dual-perspective arguments than the comparison group ( $M=1.17$  vs.  $.36$ ),  $t(86)=3.85$ ,  $p<.001$ . Percentages appear in table 3.

*Questions.* Table 4 shows mean number of questions posed at Y3. Experimental group means exceeded comparison group means for both topics - for TP,  $t(69)= 7.63$ ,  $p < .001$ ; for euthanasia,  $t(86)=3.32$ ,  $p=.001$  . Moreover, questions of a type we labeled case-based were more prevalent in the comparison groups. Illustrations of each type appear in table 5.

## **Conclusions:**

We believe these results are significant on a number of grounds. There is much talk in education circles about the importance of “21<sup>st</sup> century skills.” But these can have weight only to the extent that they can be rigorously defined and measured. The present work shows that the skills we focus on can be identified and assessed with precision, and, critically important, can be developed. In so doing it also shows that empirical research can contribute to articulating educational goals (rather than only means of achieving goals stipulated by others).

Acknowledgement of the value of these skills is essential if the method presented here is to be regarded as justified, given its extended time investment and “stand-alone” (non-curriculum-embedded) nature. In the forms examined here, the skills we identify are not terribly high-level cognitive skills. Yet they arguably are fundamental to the kinds of higher-order thinking 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 indication that the skills identified here develop naturally during the age range examined. The proportion of participants showing dual-perspective argument remained steady at about a third.

The method of developing these skills presented here has a number of positive attributes. Dialogic argumentation skill has not been a major concern of educators, and we focus here instead on evidence regarding the kinds of individual performance that has warranted much more attention in educational circles. Yet the development of 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 remain a significant educational challenge.

An additional component of the outcome of the intervention reported here warrants highlighting – the epistemological. The data on student-posed questions, we believe, indicate that students acquired not simply a question-asking routine, or habit, but rather an awareness that evidence was relevant to their arguments, especially arguments about social issues, compared to science topics where a prevailing “science-as-accumulated-fact” epistemological stance may make this awareness less challenging to acquire (Kuhn, 2010).

A complex, multi-component intervention of course requires further experimental dissection so as to isolate its effective components, work that remains to be done. In the absence of compelling evidence of success of more “quick fix” approaches, the effort appears worth pursuing. Approaches that arguably have greater face validity, such as the extensive practice in essay writing engaged in by our comparison groups, appear not as effective as a less direct approach.

## Appendices

*Not included in page count.*

### Appendix A. References

*References are to be in APA version 6 format.*

- Common Core State Standards Initiative. (2010). [www.corestandards.org](http://www.corestandards.org).
- Graff, G. (2003). *Clueless in academe: How schooling obscures the life of the mind*. New Haven: Yale University Press.
- Kuhn, D. (1991). *The skills of argument*. Cambridge: Cambridge University Press.
- Kuhn, D. (2010). Teaching and learning science as argument. *Science Education*.  
(*Published Online: Mar26 2010 DOI: 10.1002/sce.2039*).
- 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 D.C. : U.S. Dept. 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*. New York: Cambridge University Press.
- Shaenfield, D. (2010). *Arguing with peers: Examining two kinds of discourse and their cognitive benefits*. Manuscript submitted for publication.
- Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge MA: Harvard University Press.

**Appendix B. Tables and Figures**

*Not included in page count.*

**Table 1**  
**Coding Scheme for Teacher Pay Essay**

Argument Type	Examples	
	Equal Pay is Preferred Option	Experience-based pay is Preferred Option
<b>No argument</b>	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.
<b>Own-side only</b> (Includes only Positives of Preferred Option)	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.	Teachers with more experience should get paid more because they are the ones that worked hard to get in the position they're in.
	I think all teachers should get the same pay because think how hard ALL teachers work.	Experienced teachers should get more pay because of the skills they have, because they have more to offer the students.
<b>Dual Perspective</b> (Includes Negatives of Other Option)	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	If new teachers got the same pay, experienced teachers would get fed up and quit.
	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.	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.
	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?	
<b>Integrative Perspective</b> (Includes Negatives of Preferred Option or Positives of Other Option)	Experienced-based pay may seem fair to those who have taught for a long time [positive other]. But not for the new teachers who do just as much as everyone else [negative other].	Although it does seem unfair the school is basing your salary on age [negative own], it's a clever way to keep good teachers for a longer time [positive own].
		This cannot be viewed as unfair to new teachers [negative own] because with more experience they will receive the pay they deserve [positive own].

Table 2

Percentages of Participants Making Dual-perspective & Integrative Arguments  
by Time & Condition (TP essay)

	<b>Initial</b>		<b>Year 1</b>		<b>Year 2</b>		<b>Year 3</b>	
	<b>E</b>	<b>C</b>	<b>E</b>	<b>C</b>	<b>E</b>	<b>C</b>	<b>E</b>	<b>C</b>
Percentage of participants making any dual-perspective arguments	35	35	67	38	79	19	79	29
Percentage of participants making any integrative arguments	00	00	00	00	00	00	30	00

Note. E=Experimental; C=Comparison.

Table 3

Percentages of dual-perspective and integrative arguments at Y3  
(Posttest-only Euthanasia Essay)

<b>Dual-Perspective Arguments</b>			
	<b>Experimental</b>	<b>Comparison</b>	<b>External Comparison</b>
Percentage of participants making any dual-perspective arguments	<b>73%</b>	<b>29%</b>	<b>24%</b>
<b>Examples</b>			
(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.			
<b>Integrative Arguments</b>			
	<b>Experimental</b>	<b>Comparison</b>	<b>External Comparison</b>
Percentage of participants making any integrative arguments	<b>30%</b>	<b>18%</b>	<b>08%</b>
<b>Examples</b>			
(Pro-Euthanasia) Even though it might not be easy to stand by and watch [negative own], a family member must support whatever the ill person wants [positive own].			
(Anti-Euthanasia) Sure it will take away the pain right away [positive other], but that wouldn't be the only thing gone [negative other].			

Note. The post-only assessment in year 3 allowed inclusion of students not included in the longitudinal sample due to missing assessments, either because of absence or because they did not enter the school until between the middle of year 1 and beginning of year 3 (almost all entered at the beginning of year 2), increasing sample size to 60 experimental (31 female) and 28 (13 female) comparison. 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.

Table 4  
Mean Number of Questions Posed

	<b>Teacher Pay Topic</b>	<b>Euthanasia Topic</b>
<b>Y3 Experimental</b>	4.18 (10% case-based)	3.26 (0% case-based)
<b>Y3 Comparison</b>	.33 (100% case-based)	1.64 (74% case-based)
<b>Y3 External</b>	NA	.92 (30% case-based)

Table 5

## Questions Participants Posed about Euthanasia

## General Questions

How many people in a hospital (average) have an incurable disease?  
 What incurable diseases are there and what are the chances of surviving them?  
 What is considered an incurable illness?  
 What are the most painful diseases?  
 What are the treatment options for incurable diseases (e.g. cancer)?  
 What can doctors do to ease pain for someone?  
 How many people that have incurable illness go to a therapist?  
 Does depression affect suicide?  
 What percent of people who have an incurable disease want to die?  
 How many people have asked doctors to let them die?  
 How many people are assisted with killing themselves?  
 How would a life-ending procedure go? For example, would it be a shot?  
 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?  
 How are families included in the deaths of patients? How do they assist them?  
 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?  
 How many states have made a law in favor or against assistance deaths?  
 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?

## Case-Based Questions

Who is the person? (Adult, child, etc)  
 What is their disease?  
 How old is the patient?  
 In what conditions is this person in?  
 How long do they have to live?  
 How long the person can live with the disease?  
 Is there any way to cure them?

Are they in pain?

How long do they have to live?

Does the family care what their decision is?

Are they on life support?

What are their symptoms?

Do they have a family – for example if the women had kids?

Where do they work?

What do they do for a living - for example if they have an important role in society?

Note. Eliminating very similar questions, these are all questions that were asked. They were not answered (in contrast to questions asked during the intervention).

Table 6  
 Percentages of Participants Making Dual-perspective & Integrative Arguments in Replication Sample

	Experimental group			Comparison group
	Initial	Year 1	Year 2	Year 2
Percentage of participants making any dual-perspective arguments	30	54	73	38
Percentage of participants making any integrative arguments	00	03	14	00

Note. Sample size is 37 experimental and 21 comparison. Two experimental participants are missing Y0 data and two Y1 data.

Figure 1  
Mean Number of Dual-perspective Arguments on Teacher Pay Essay

