Dialog as a Bridge to Argumentive Writing

Yuchen Shi, Flora Matos & Deanna Kuhn

Teachers College Columbia University, New York, NY | USA

Abstract: We describe a dialogic approach to developing argumentive writing whose key components are deep engagement with the topic and extended discourse with peers that provides the activity with both an audience and a purpose. In a dialogic intervention extended over an entire school year, pairs of sixth graders engaged in electronic discourse with peers on a sequence of topics, as well as wrote individual final essays on each topic. In their essays, they showed achievements relative to a non-participating group in coordinating evidence with claims, in particular in drawing on evidence to weaken claims as well as to support them. They also showed some meta-level enhancement in understanding of the role of evidence in argument. A recall task ruled out the possibility that this enhancement was due to superior recall of the specific evidence available to them, rather than broader meta-level understanding. A case is made for fostering development in argumentive writing both dialogically and in the context of topics that students engage with deeply.

Keywords: dialog, argumentation, evidence, writing

journal of WRITING RESEARCH

Shi, Y., Matos, F., & Kuhn, D. (2019). Dialog as a Bridge to Argumentive Writing. *Journal of Writing Research, 11*(1), 107-129. https://doi.org/10.17239/jowr-2019.11.01.04

Contact: Flora Matos, Teachers College Columbia University, 525 W 120th St New York, NY 10027 | USA - flora.matos@tc.columbia.edu.

Non-narrative argumentive writing is a staple of academic success. Central to argumentive writing is the coordination of claims and evidence and the use of evidence-based claims to strengthen one's own position and to weaken that of an opponent. Needed ideas are not all initially available and they must be organized into a more complex form than a simple linear one. Argumentive writing thus poses greater cognitive demands than narrative writing and students of all ages find it difficult. In fact, students' struggles to do so at all educational levels have been well documented and analyzed, along with varying kinds of approaches for remedying them (Graham, & Perin, 2007; Newell, Beach, Smith, & VanDerHeide, 2011; Ferretti & Lewis, 2013). These approaches for the most part work directly on students' writing and revision. Here we report on an approach to supporting students' argumentive writing development by building on its developmental roots, in other words by employing dialog as a bridge from the conversational exchanges that come naturally to children to the individual written production that does not. In what follows, we make a case for its power.

The dialogic approach represented in the present work can be traced as far back as Socrates, but in contemporary empirical research it fits most closely within a sociocultural framework and its core idea of collaborative cognition, or thinking as a social practice (Cole, 1998; Tomasello, 1999). Philosophically, it is most closely aligned with the work of Walton (2014) and van Eemeren and Grootendorst (1992), who regard it as essential to evaluate arguments within a dialogic context.

The strength of a dialogic approach, Graff (2003) has proposed, resides in its providing the otherwise absent interlocutor. The student writer stares forlornly at the blank page, hoping to somehow fill it with bland statements at least vaguely related to the assigned topic but directed to no one, without saying anything anyone might object to. In dialog, the student knows who he or she is addressing and has a purpose in doing so. These two components – a clearly defined audience and a meaningful purpose – are essential to effective writing. Without them, student writers are at risk of struggling simply to fill the page with "what the teacher wants," which the students take to mean statements loosely related to the topic that do not risk stating anything that might be questioned (Graff, 2003).

A number of contemporary education theorists have championed the educational power of discourse, among the most notable Resnick and Mercer and their respective colleagues (Resnick, Michaels, & O'Connor, 2010; Resnick, Asterhan & Clarke, 2015; Mercer & Littleton, 2007). They highlight the value of discourse engagement and practice in its own right, with the implications for individual writing mostly only implied. Researchers who have followed their dialogic lead most often have focused their investigations at the whole-classroom level of discourse, although Reznitskaya et al. (2001) and Nussbaum and colleagues (Nussbaum & Edwards, 2011) have also studied children working in small groups. Our own approach, reflected in the present work, emphasizes dyadic discourse, i.e., between two parties who speak or write directly to one another and are directly responsible for maintaining the exchange. Its

strength, in our view, is its maximizing of cognitive engagement, i.e., the individual is constantly on call to respond to the other and maintain the continuing exchange. It also effectively removes the teacher as the center of a wheel through which all interaction is channeled.

1. Developing Argumentive Writing through Discourse

Why might dialog serve as an effective bridge to individual writing? Instead of beginning by working directly on pieces of writing students produce, the core idea underlying dialog as a bridge to individual writing is Vygotsky's (1978) idea of knowledge transmission from the inter-individual to the intra-individual level, after sufficient exercise at the inter-individual level. Intermediate reflective writing (e.g., the Vee-diagrams introduced by Nussbaum, 2008, or the reflection sheets in the present work) serves to strengthen the bridge. At the same time, we agree with others who have sought to integrate discourse with individual writing, and one or both with disciplinary knowledge acquisition (DeLaPaz, Monte-Sano, & Felton, 2017; Chen, Park, & Hand, 2016), in maintaining that a rotation across individual and social modes, rather than emphasizing one to the exclusion of others, is the most promising approach.

In addition to capitalizing on its developmental roots, a further benefit dialogic argument offers is its fostering of what Nussbaum and Asterhan (2016) refer to as *proactive executive control strategies*. What am I undertaking to accomplish and how am I undertaking to do so? We can call this "meta-strategic awareness," and later we describe our effort to assess it, but the point to make now is that it is as fully necessary in writing as it is in discourse. This is especially the case in individual writing, as there is no external other to assist in such executive control functions (Zillmer & Kuhn, 2018).

Dialog as a path to argumentive writing, we have suggested, provides two components critical to successful writing – a clearly defined audience and a meaningful purpose. Toward this end, participants choose the position they will take. A core feature of the dialogic method employed in the present work is that it is student-centered, emphasizing peer-to-peer rather than teacher-centered interaction. Underlying it is a commitment to the view that high-level intellectual skills such as argumentation are sufficiently important to warrant dedicated attention in their own right as curriculum goals.

Further, our approach is based on the view that developing argumentation skills, and the values to support them, necessitates sustained and dense practice in rich environments that require those skills and values. This engagement entails both a supportive community and the strengthening of individual skills and understanding. Hence it is not quickly achieved. In the work described here, students engage deeply with a series of challenging argumentive topics, over an extended period. They also engage in rich discourse of two forms, the first one verbal with a same-side peer to prepare for dialogs with successive opposing-side peers that follow. This essentially

doubles the participation in reasoned discourse, with both types encouraging metacognitive planning and reflection.

A further characteristic of the present approach is that discourse is conducted verbally (with a same-side partner) but also electronically, between a same-side pair and a succession of opposing pairs. Doing so provides a written document that externalizes thought into a tangible form, in contrast to verbal discourse, which disappears as soon as uttered. The electronic medium thus facilitates reflection on what is exchanged, taking discourse temporarily "off line" (Olson & Oatley, 2014). In addition to serving as a reference point and framework during discourse, these transcripts become the object of various reflective activities students engage in.

2. The Role of Evidence in Argument

A core component of argumentation, as noted at the outset, is an evidence-based claim, requiring the coordination of a claim with evidence bearing on it. This is the basic unit we have used previously (Kuhn, Hemberger, & Khait, 2016a,b) and use in the present work in examining students' argumentive writing. Increasingly, researchers in the field of science education, as well as that of argumentation more broadly, have come to focus on the use of evidence in argument (Asterhan & Schwarz, 2016; Chen et al., 2016; Kuhn, 2018; Kuhn & Moore, 2015; Manz & Renga, 2017; McNeill & Berland, 2016). Coordinating claims and evidence makes different kinds of demands. In particular, drawing on evidence to weaken a claim appears to be a more challenging achievement than does using evidence to support a claim (Kuhn & Moore, 2015).

How are students to gain access to the evidence they need in order to coordinate claims and evidence and hence to argue well? A possible drawback of the traditional method of asking students to begin their work on a topic by reading texts on it is that students do not yet appreciate the purpose this information might serve. In a word, it provides answers to questions they do not yet have. As a result, they fail to see its point. They are thus likely to approach such reading disinterestedly, as just another reading assignment to complete. An alternative method we have devised is to enable students to first see how such information could be useful by providing them the opportunity to pose their own questions, in addition to factual information we provide in brief Q&A format. In so doing, we help them to situate knowledge about the topic in a framework of the arguments it has the potential to serve.

3. Tracing the Evolution of Argumentive Writing

The method employed here has been shown to be effective in developing skill in both contexts (Kuhn & Crowell, 2011; Crowell & Kuhn, 2014; Kuhn & Moore, 2015; Kuhn, Hemberber & Khait, 2016a,b). From a research perspective, our approach has the advantage of allowing us to observe a progression as new skills develop in connected ways in both the dialogic and individual writing contexts.

During their engagement with a succession of topics, we have closely followed the progress students make in the quality of their final individual essays on each topic and observed a developmental progression, in particular with respect to the nature of the evidence writers draw on to justify their claims (Kuhn et al., 2016a,b). Although evidence of multiple types has been generated and is available to them, students initially confine themselves largely to use of evidence to support their own claims (upper left box in Figure 1). Only later do most students begin to use evidence to weaken the opposing claim (lower right box in Figure 1). As represented by the diagonal connecting line in Figure 1, these forms of evidence work together to serve a dual argumentive strategy: "Here's everything good about my position and bad about yours." Later still do some students begin to cite either of the remaining two types of evidence in Figure 1 (typically "Support Other" and rarely "Weaken My") – evidence that cannot as readily be used to support their own position and rather, if noted at all, needs in some way to be reconciled with it. This is difficult to do, and writers tended to initially simply note and quickly ignore such evidence – "It's true that _____, but anyway . . ." – only at a later stage attempting an integration in a genuine "However" structure that connected two adjacent of the four types in Figure 1.

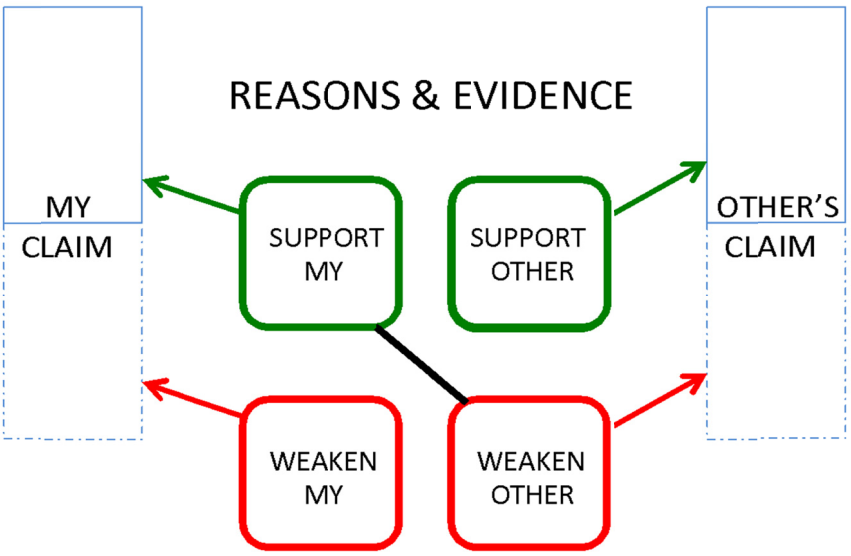

Figure 1. Forms of coordination of claims and evidence.

Although reassuring that progress is achievable, this is slow, labor intensive and incomplete progress. In our most recent work (Hemberger et al., 2017), we therefore explored the extent to which it is possible to accelerate it. We undertook to do so by providing students a further degree of support in developing their use of evidence of

different kinds to bear on claims. We maintained the brief Q&A format for evidence – students are encouraged to generate their own questions that they are provided answers to; in addition, however, we provided students one carefully selected piece of evidence (also in Q&A format) during each of their dialog sessions, with a prompt to "try to say something about this evidence in your dialog today."

Evidence of different types was presented to an experimental group of sixth graders in what we hypothesized to be an optimal sequence, based on their observed order of emergence in the earlier work (Kuhn et al., 2016a). Students of course often heard evidence supportive of the opposing position directly from their opponents during the dialogs. We thus asked whether a prompt to try to address a piece of evidence supporting the opposing position would be beneficial.

This effort met with success in accelerating the experimental group's use of evidence-based claims in both dialogs and essays over the course of a year-long intervention, relative to comparison groups who were presented no additional evidence or only evidence supporting their position. By the final topic essay at the end of the year, the experimental group showed significantly more use of evidence in their essays (a mean of 3.16 pieces, compared to .83 at the first topic essay) than either comparison group, thus transferring their newly developing skill from one topic to another (Hemberger et al., 2017). Use of different types of evidence emerged in a sequence corresponding to the cognitive demands they posed. Students first used Support-own evidence. They used Weaken-other evidence increasingly over time, but the two evidence types inconsistent with their position (Support-other and Weaken-own) showed lesser and later gains. Superiority of the experimental group over the two comparison groups establishes that it is not simply availability of evidence per se that provides this benefit. Finally, and further supporting a dialogic approach, qualitative data showed evidence use occurred most readily in dialogs; only more gradually did it appear in individual writing on the same topic, and to a more limited extent in essays on a new, unstudied topic.

4. The Present Study

The intervention method employed in the present study replicates, with a new group of sixth grade students, the method used by Hemberger et al. (2017). A central component of the curriculum was same-side pairs of students' engagement in electronic dialogues with a series of opposing-side pairs. During these e-dialogues, we encouraged students to generate questions on the topic that they wished answered and answers were provided at the next class session; we made available additional pieces of evidence in a Q&A format. Evidence serving all four argumentive functions was made available to students in what we identified as an optimal sequence: support own position, weaken opposing position, support opposing position, and weaken own position.

We add to this study some additional measures, with a goal of investigating students' understandings regarding evidence connected to a claim as playing a key role

in argument. By administering these measures to both experimental and comparison groups, we sought to explore to what extent participating students come to recognize the key role of evidence in argumentive discourse and writing. Apart from skill development, does sustained engagement in the intervention enhance students' meta-level understanding of the purpose and goals of evidence in argumentive writing?

We hypothesize that sustained practice in seeking and using evidence in argumentation should promote students' meta-strategic awareness in this respect, by creating a need for use of evidence in the service of argument, to both support own-side and weaken opposing-side claims, and in so doing to better appreciate its value. Earlier work tracing middle-school students' meta-level statements to one another as they engaged in the dialogic intervention showed growth over time in this meta-level understanding of argumentive discourse (Kuhn et al., 2013), leading us to ask to what extent this same evolution would be evident in students' writing. For this purpose, we examine students' prior selection and subsequent recall of evidence, in addition to its use in their essays. In so doing, we see this thinking *about* evidence (in contrast merely to use of it) as reflecting their meta-level understanding of it and its function (Kuhn, 2001). Our main hypothesis is that we will see advancement in this respect over the course of students' engagement, as indexed by comparison with a group who did not participate in the intervention. Specifically, will they recognize the relevance of different kinds of evidence beforehand in contemplating their writing task and will this perceived relevance lead them to better represent and therefore recall it later? Based on previous findings (Hemberger et al., 2017), we can also predict growth at the strategic level, i.e., in students' use of different kinds of evidence-based claims in their argumentive writing itself.

5. Method

5.1 Participants

Participants were 54 sixth-graders (all 11 or 12 years old) enrolled in an urban public middle school in an underserved neighborhood of a large Northeastern city in the United States. Participants were from similar ethnic, socioeconomic, and academic backgrounds, evenly divided by gender, 96% of Hispanic or African-American ethnicity and 60% eligible for free lunch. The majority were functioning below grade level and regarded as academically at risk.

All entering students were assigned randomly to three sixth-grade classes by the school administration at the beginning of the school year. On the basis of this random assignment of students all new to this middle school, we regarded classes as equivalent groups, and the administration confirmed that in addition to demographic equivalence, academic performance of the classes began and remained at comparable levels.

5.2 Design

Two classes chosen randomly participated in the study. One class served as the comparison condition and took part only in a final assessment, equivalent to and administered at the same time as that administered to the experimental group. The other class served in the experimental condition and participated in the curriculum on a twice-weekly basis for the entire school year, prior to the year-end final assessment. While the experimental group participated in the intervention, the comparison group engaged in their regular Social Studies classes, which did not include debate or argumentation or full essay writing.

Of the 54 students who entered 6th grade at the beginning of the study, 49 remained in the final sample. In the experimental condition, of the 27 students who began the intervention, 22 (13 females) remained in the final sample. Five students from the experimental condition were excluded due to excessive absence (more than 50% of intervention sessions). In the comparison condition, 27 students (13 females) participated on a single occasion in an assessment conducted at year's end that was identical to the assessment conducted with the experimental group.

5.3 Intervention Procedure

The intervention procedure consisted of a year-long dialog-focused argument curriculum closely replicating the intervention implemented in previous studies and reported by Hemberger et al. (2017). (For full details of the intervention procedure, see Kuhn et al., 2016b.) The intervention was divided into four cycles each consisting of 13 twice-weekly 40-minute class sessions, with a new topic introduced at the beginning of each cycle. From an initial set of 10 topics polled prior to the start of the intervention, four topics were chosen for the intervention as ones for which students' opinions most closely approximating an equal split of students favoring pro and con sides. Topic 1 was whether a tax should be imposed on purchase of soft drinks. Topic 2 was whether a parent arriving from a foreign country should be permitted to home school their child. Topic 3 was whether the U.S. should assist a South American country invaded by a neighboring country. The final topic was whether teens should go directly to college after high school or work for a period first. Each topic cycle started with same-side group work (*Pregame* sessions), followed by paired electronic dialogues with a series of opposing-side pairs (*Game* sessions). Final same-side group work (*Endgame* sessions) and finally a whole-class *Showdown* debate. The topic cycle concluded with a debrief session and a final individual essay assignment. The intervention was implemented by the authors, who had designed and had previous experience implementing this intervention with middle-school low-achieving populations.

Pregame (Sessions 1 and 2)

After choosing their position on the topic, students gathered in same-side groups of three to five students. An adult coach facilitated discussion if needed. In the first

session, *Our Reasons*, students generated reasons as to why their favored position was the better one. These supporting reasons were written on cards, one reason per card, shared with peers, and rewritten for clarity as needed. In the second session, *Evaluating Reasons*, students further discussed the reasons and collaboratively sorted them reason cards into three categories: best, good and so-so.

Game (Sessions 3 to 8)

Students formed same-side pairs that remained intact throughout this phase. Using simple word-processing software, a pair engaged in an electronic dialog with a different opposing-side pair at each session. In addition to collaborating with the partner in deciding on the dialog input, the pair worked together on an own-side or other-side *Reflection Sheet,* while awaiting the opponents' electronic response. These prompted students to reflect either on one of their arguments or on one of their opponents' arguments, and on the best counterarguments and rebuttals in each case.

During Topics 1 to 3, students were also encouraged to construct evidence questions, the answers to which they thought might help them in making their arguments. These were answered for students by the next session and shared with the group. As a supplement to these questions, students were also provided two to five (depending on topic) pieces of potential evidence in Q&A format by the conclusion of the *Game* segment. Doing so insured that students by the end of their work on the topic had encountered evidence that embraced all four argumentive functions (support own, weaken other, support other, weaken own).

Endgame (Sessions 9 to 13)

Students returned to their same-side groups in order to prepare for the whole-class *Showdown*. At one session, a *Summary Reflection Sheet was* completed to facilitate students' review of the other side's arguments and their counterarguments against them. During the following session, the group prepared a second summary sheet that reviewed their own arguments, expected counterarguments and rebuttals, and discussed strategies for the *Showdown*.

During the whole-class *Showdown* session, a succession of students from each side volunteered to verbally debate a classmate from the opposing side. During this three-minute period either debater or any of their teammates could call a one-minute *Huddle* to enable the speaker to receive help from teammates. These verbal exchanges were video recorded and transcribed for the preparation of an argument map to be used in the *Debrief* session. During this session, students were guided through the argument map, where points were assigned for effective argument moves (e.g., counterarguments, rebuttals, and evidence use) and subtracted for ineffective moves (e.g., unwarranted assumptions, unsupported claims, and misuse of evidence). A winning side was declared based on these points.

To conclude work on the topic, students wrote an individual final essay in the form of a position statement written as a newspaper Letter to the Editor. Students completed

these in a whole-class setting, and were given an entire class period (40 minutes) to complete them. Most finished within 30 minutes and all in 40 minutes. A copy of the complete Q&A-format list of evidence on the topic that had been made available earlier remained available for students' use while they wrote the essay. All students received the same verbal prompt: "Here is some information relevant to the topic; remember not all information is going to support the side you favor. If it doesn't, see if you can deal with it. Feel free to use any information you would like in your essay, but you are not required to do so." Simple clarification of word meaning or task instruction was provided if requested. Students were told there was no length requirement for the essay.

The next session marked the beginning of a new topic and the sequence of activities was repeated. The entire sequence of the topic cycle is summarized in Figure 2.

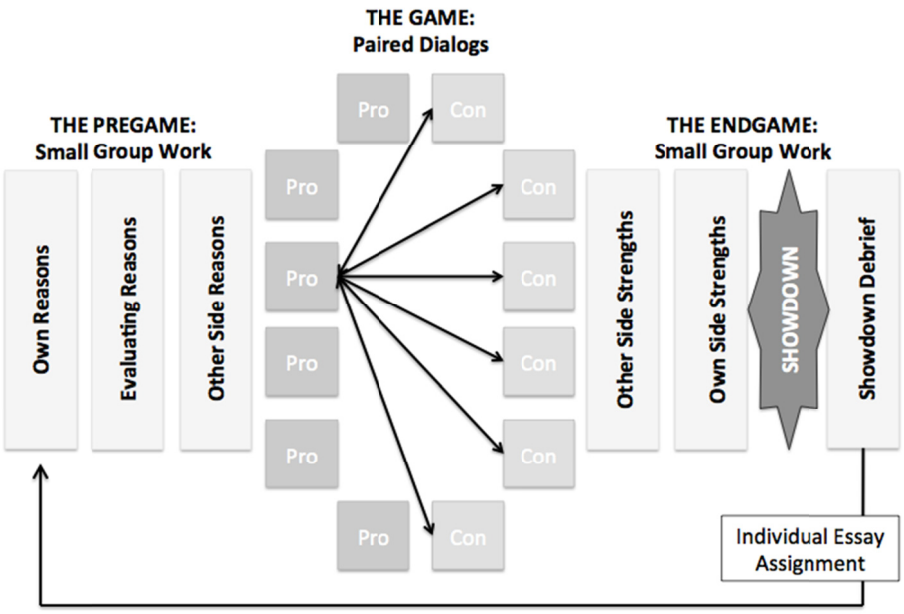

Figure 2. Topic workflow from pre-game to final essay (From Kuhn et al., 2013).

5.4 Intervention Topic Assessment

In addition to the main essay component of the post-intervention assessment, there occurred a pre-essay and a post-essay component involving evidence selection and evidence recall. The assessment was linked to the experimental group's fourth topic (college vs. work), as we wanted intervention students to write on a topic they had engaged with deeply (a condition central to the intervention method). However, since the focus of the assessment was on use and understanding of evidence, we postponed evidence generation during the experimental group procedure for Topic 4 until this

final session, in order to equate exposure to evidence items across experimental and comparison groups. The experimental group also wrote an essay on a new topic at the following class session to assess transfer of skills to an unstudied topic.

Both intervention and comparison groups were told they would write individual essays on the college vs. work topic and complete several related tasks. As elaborated in the report of results, essays were coded based on writers' success in making functional evidence-based claims, defined as claims that make a clear connection between evidence and claim. The other tasks were included as a means of examining the hypotheses indicated in the introduction regarding students' understanding of the nature and relevance of evidence.

Evidence selection

As the first task, students each received a written list of four types of evidence they could potentially use when writing the essay:

1. Evidence of good results that come from going right to college
2. Evidence of good results that come from working before going to college
3. Evidence of bad results that come from going right to college
4. Evidence of bad results that come from working before going to college

The written instructions asked students to circle the type of evidence they would most like to see before writing the essay. They were then asked to circle the type they would second most like to see.

In the next part of the task, students were asked to choose specific evidence questions they would like to have answers to prior to writing the essay. Twelve questions, without answers, were listed in the same order as questions and answers in the complete Q&A evidence list they would see a moment later (see Appendix). Students were encouraged to choose as many questions as they would like from the list.

Essay writing

Students then proceeded to the essay. The essay prompt instructed students to write a letter to other students in their last year of high school, and to give them advice on whether it's better to go right to college or work first. Students were told the goal of the letter was to change as many students' minds as they could so as to be consistent with the student's own view.

Before students began writing, the coach distributed to each of them the list of 12 pieces of evidence in Q&A format (see Appendix). These were balanced so as to serve four potential argumentive functions - support own position, weaken other's position, support other's position, and weaken own position. Students were told to feel free to use the evidence provided when writing the essay but that it was not required to do so. The majority of students finished their essays within a 20-min period, but the entire 40-min class period was available to them.

Evidence recall

Essays and evidence sheets were collected before students proceeded to the next task. They received a sheet containing the list of 12 evidence Qs (see Appendix), but now without answers. They were asked to recall the answers to these questions. They were told it was fine to write the general idea if they didn't remember the exact answer. Most students completed the task within a 10-min period.

5.5 Transfer Topic Assessment

To compare the writing performance of intervention students on a studied and unstudied topic, at the next class session intervention students were asked to write an individual essay on the topic of whether teens who commit serious crimes should be tried in an adult court system or a juvenile court system. Instructions were identical to those for the main assessment topic. They were provided a similar list of 12 pieces of Q&A evidence to use in their essays if they wished.

5.6 Coding of Essays

Each essay was first segmented into idea units, with an idea unit defined as a claim together with any reason and/or evidence supporting it. The next step involved assigning each unit as either evidence-based or non-evidence-based categories. Given the focus of the present investigation on the use of evidence in argument, only evidence-based units were further analyzed. Following Hemberger et al. (2017), these were further categorized as functional vs. non-functional. A claim can be regarded as evidence-based only if the function of the evidence is made clear by linking it to the claim it is intended to serve. If the connection between evidence and claim was missing (e.g., the evidence was merely stated with no implication drawn) or the evidence was mischaracterized, the unit was coded as non-functional and not coded further. Functional evidence-based claims were further classified into one of four categories based on their specific function: support own side, weaken opposing side, support opposing side, and weaken own side.

Two of the authors randomly chose 20% of the dataset and independently segmented them into idea units, achieving an inter-rater agreement of 93%. Having resolved disagreements in segmentation through discussion, they proceeded with assigning each unit into one of the six categories (non-evidence-based category, non-functional evidence-based category, four functional evidence-based categories), achieving an agreement of 83% (Cohen's kappa = 0.736, $p < 0.001$). Disagreements were settled through discussion and one of the authors coded the remaining essays.

Each level is defined and illustrated in Table 1, which refers to the following piece of evidence:

Q: Do most good jobs require college?

A: Yes. It is estimated that by year 2020, 35% of all jobs will require at least college education. High paying jobs such as those in science and engineering always require at least a college degree.

Table 1. Levels and Examples of Coding of Evidence-based Essay Units

Level	Category	Example	Writer's position
Functional evidence-based claims	*Support-own*: an evidence statement serving to functionally support one's own position.	It can help you get a better job, more pay, and you will learn stuff along the way. By 2020, 35% of jobs will require at least a college degree to get it.	College
	Weaken-other: an evidence statement serving to functionally weakens the opponent's position.	Good jobs like in fields of science and engineering require at least a college degree. This means that if you have a passion for science or engineering you won't be able to pursue your dream without a college degree.	College
	Support-other: an evidence statement serving to functionally support the opponent's position.	However, some people say that you should go to college first because with a college diploma you get more money.	Work
	Weaken-own: an evidence statement serving to functionally weaken one's own position.	However, if you work for one year before going to college you and your parents don't have to worry about the expenses. You might wonder if you can get a great job while being in high school.	Work
Non-functional evidence-based claims	Attempted use of evidence to justify a claim without a discernible connection between evidence and claim.	I want to change because what if you don't have a high school diploma you have to get a job. Like by year 2020, 35% of all jobs will require at least college education.	College
	Simple re-statement of evidence unconnected to any claim. [Can be a full or partial verbatim copy of evidence or a reasonably accurate paraphrase of evidence]	And also yes, it is estimated that by year 2020, 35% of all jobs will require at least college education. High paying jobs such as those in science and engineering always require at least a college degree.	College
	Evidence is mischaracterized and cited in a way that substantially misrepresents its meaning.	No most jobs don't require college because they need people who work hard and have special skills.	Work

6. Results

6.1 Intervention Topic Assessment

Essay writing

We first examined whether the length of the essays differed significantly across the two conditions by the number of idea units an essay contains. The mean number of units was 6.09 (*SD* = 3.60) for the experimental condition and 4.93 (*SD* = 3.31) for the comparison condition. A Generalized Linear Model (GLM) using the Poisson distribution indicated the length of experimental essays was 1.237 times the length of comparison essays, a non-significant difference, *Wald* $X^2(1, N = 49) = 3.008$, $p = 0.083$.

Given the focus of the present investigation is students' use of evidence, in subsequent analyses we focus only on evidence-based units. We first examine whether the two conditions differed in mean number of evidence-based units. These means were 3.00 (*SD* = 1.66) for the experimental condition and 2.67 (*SD* = 2.34) for the comparison condition. The number of evidence-based claims in the experimental condition is 1.125 times that in the comparison condition, a non-significant difference, *Wald* $X^2(1, N = 49) = 0.478$, $p = 0.489$.

A significant difference between groups was found, however, in the number of functional evidence-based claims. The experimental group made an average of 2.68 (*SD* = 1.89) functional evidence-based claims, while the comparison group made an average of 1.30 (*SD* = 1.46) functional evidence-based claims. The number of evidence-based claims made by the experimental condition was 2.069 times that of comparison condition, a significant difference, *Wald* $X^2(1, N = 49) = 11.610$, $p = 0.001$. Thus, the students in the experimental group were more successful in their use of evidence in their essays.

Functional evidence-based claims were further examined by category – support-own claims, weaken-own claims, support-other claims, and weaken-other claims, as shown in Table 2.

The experimental group was significantly more successful in making both support-own and weaken-other evidence-based claims. The number of support-own functional evidence-based claims made by the experimental condition was 2.455 times that of the comparison condition – a significant difference according to the GLM using a Poisson distribution, *Wald* $X^2(1, N = 49) = 8.063$, $p = 0.005$. Similarly, for weaken-other functional evidence-based claims, the mean was 1.18 (*SD* =1.14) for the experimental group and 0.59 (*SD* = 1.12) for the comparison group, a significant difference. The number of weaken-other evidence-based claims made by the experimental group was 1.994 times that of the comparison group, *Wald* $X^2(1, N = 49) = 4.720$, $p = 0.030$.

These results indicate that the dialog-focused argument curriculum not only strengthened students' ability to successfully use evidence to support claims in favor of their own side; it also enhanced their ability to use evidence to counter the opposing side, a more demanding skill commonly neglected by novice writers.

Table 2. Means (and Standard Deviations) of Four Types of Functional Evidence-based Claims in Essays by Condition

Types of evidence-based claims	Experimental condition (n=22)	Comparison condition (n= 27)	Exp(B)
Support-own	1.36 (1.36)	.56 (.70)	2.455**
Weaken-other	1.18(1.14)	.59(1.12)	1.994*
Support-other	.091 (.29)	.15 (.46)	.614
Weaken-own	.05(.21)	.00(.00)	.001
Total	2.68 (1.89)	1.30 (1.46)	2.069**

Note. *p< 0.05, **p<0.01

As we anticipated, however, the successful use of evidence in the essays of both groups was largely limited to these two functions. Students in neither group often made reference to evidence that would support the opposing position or weaken their own position. These two types of evidence use are most cognitively challenging since they are inconsistent with students' own position. The use of both of these two types of evidence-based claims was negligible as reported in Table 2 and not statistically significantly different across groups.

Students' successful use of evidence of the two less challenging types was not as frequent as it could have been, given the evidence they had available to them – only slightly above a mean of one such use in the experimental condition and below one in the comparison condition. We therefore examined what proportion of students ever made such successful evidence claims in their essays, with the aim of assessing to what extent the curriculum benefitted all students versus only a more able few responsible for these successes. As seen in Figure 3, the majority of the experimental group – about two thirds – made support-own and/or weaken-other functional evidence-based claims at least once. The proportion for the comparison condition, in contrast, was below half and only one third in the case of weaken-other claims. Although the easiest, support-own group difference did not reach statistical significance, the difference for weaken-other evidence-based claims was significant, p = 0.047 (Fisher's exact test). Thus, the intervention was effective in enabling a majority of students in the experimental group to make weaken-other evidence-based claims at least once.

Evidence selection

Did participants show any meta-level understanding of evidence-based argument in anticipating the kinds of evidence that would be useful to them in writing their essay, as would be indicated by their expressing a preference to have access to one type over another? The answer to this question was yes in the respect that they rarely indicated that they would like to have access to evidence types inconsistent with their position

(i.e., support-other or weaken-own). As their first choice of evidence they would like to see, 90% of the experimental group and 85% of the comparison group chose support-own evidence as the category they would most like to see (a nonsignificant difference between groups).

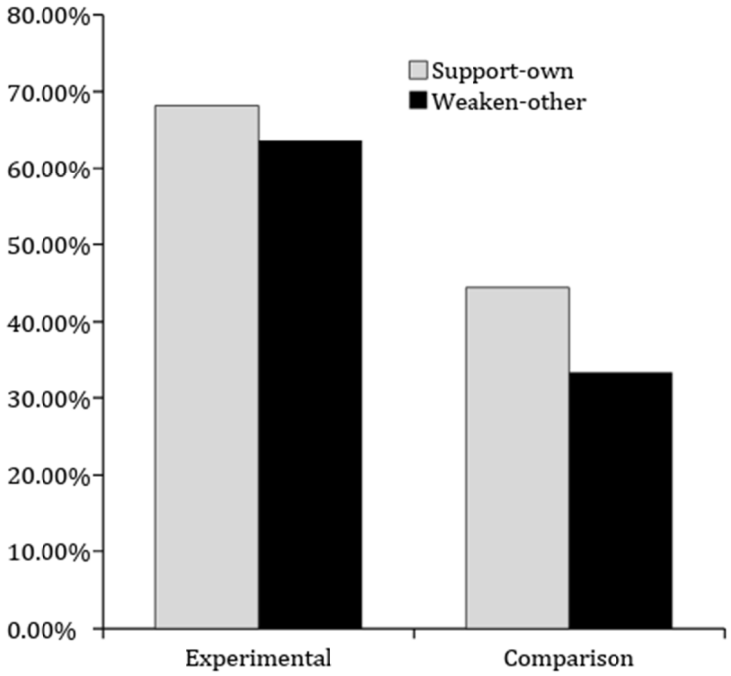

Figure 3. Percentage of students making support-own or weaken-other evidence-based claims at least once by condition

In choosing the evidence type they would like to see second most, however, a group difference appeared. Among the experimental group, 73% chose the weaken-other type, compared to only 30% of the comparison group who chose this type– a significant difference (p = .004, Fisher's exact test).

When asked to choose specific questions to receive answers to, however, this preference disappeared and no group differences appeared in selection of which questions to receive answers to (students were told that odd-numbered questions were about the work option and even-numbered questions were about the college option; only the 12 pieces of experimenter-presented evidence were included, to equalize across groups). Students in both groups tended to equally prefer evidence about the two options (in the experimental group 48.7% of selected questions were about their favored option and 51.3% about the contrasting option, a nonsignificant difference; in

the comparison group 52.6% of selected questions were about their favored option and 47.4% about the contrasting option, again a nonsignificant difference. In total, students in the experimental condition chose an average of 5.27 questions (of 12 total) and students in the comparison condition an average of 5.93 questions, a nonsignificant difference.

Were students accurate in their predictions of which answered questions they would in fact use in their essays? Neither group was highly accurate in this regard – an average of 31.5% of selected evidence appeared in the essays of students in the experimental group and 23.5% in those of comparison students, again a nonsignificant difference. (All 12 pieces of evidence were available at the time essays were written.)

Evidence recall

Participants attempted to recall the answers to most of the evidence questions – an average of 78.5% were attempted among the experimental group and 71.8% among the comparison group. Less than a third of these, however, recalled answers clearly enough to make their meaning clear, with no significant difference between groups. Among these faulty efforts, just under ten percent (9.9%) of attempts in the comparison group were faulty in recalling the evidence in a biased manner that favored their own side, and a slightly (but statistically nonsignificant) lesser proportion (7.6%) in the experimental group. (For example, in recalling the evidence "*Unemployment is high among high school graduates and research has shown that job opportunities for teens have been declining. However, usually teens can find a job working in places such as Costco, Whole Foods market, or Starbucks"*, a participant favoring the college position recalled only the first sentence and one favoring the work position recalled only the second.) In sum, then, the comparable performance by the two groups in evidence recall provides a negative answer to the question of whether any superiorities of evidence use in the essay by the experimental group are attributable to superior recall of specific evidence on the part of this group over the comparison group.

6.2 Transfer Topic Assessment

In examining performance on essays on the transfer topic, we seek to answer the question of the extent to which gains extended beyond the specific topic on which students engaged in discourse. Results for the transfer topic were compromised by a reduced N of only 13, as some students were absent due to an unannounced field trip and it was not possible given the proximity to the end of the school year to secure the missing students' data on this task. These results must therefore be treated with caution, due to the small N. Nonetheless, it is of interest to compare this group's performance on a new topic with their performance on a topic they had engaged with deeply, as well as to compare it with the performance of the comparison group also engaged with a topic they had not worked on previously. The latter comparison, however, must be

treated with caution for the additional reason that the topics the two groups were writing on were not the same (as they were in the case of the main group comparison already reported on).

Nonetheless, for the experimental group participants who completed the transfer essay, the percentage who made support-own and/or weaken-other functional evidence-based claims was 54% – only a small, and statistically nonsignificant, difference from the percentages reported in Figure 3 for the main topic. Mean frequencies of usage of these types, however, declined from 0.82 and 1.09 (for support-own and weaken-other respectively) to 0.69 and 0.54 for the transfer topic for these 13 participants. Thus, their lack of familiarity with the evidence, and the topic more broadly, negatively affected the ease with which they were able to employ the evidence. As a result, a comparison of their performance with that reported in Table 2 for the comparison group failed to show a significant difference (although, note again, the comparison is an imperfect one, as the topic differed across groups).

7. Discussion

The results reported here with respect to essay performance are comparable to those obtained by Hemberger et al. (2017), as anticipated given the two samples came from the same population, school, and grade, just one year apart in time. Both studies, as well as several preceding it using a similar method (Kuhn & Crowell, 2011; Kuhn & Moore, 2015; Kuhn et al., 2016a; Papathomas & Kuhn, 2017), support a dialogic approach to developing students' argumentive writing, especially in the population of academically low-performing students who lack experience in non-narrative writing. It is the continuing experience of dialog with a succession of peers holding the opposing position, we propose, that makes this opposing position and its accompanying arguments clear and vivid, enough so that the student can represent them in an essay and address them, and, moreover, sees the relevance of doing so**.**

The present study more specifically replicates Hemberger et al.'s (2017) finding that progress in argumentive writing can be further scaffolded by prompts that exemplify the functions of evidence in relation to a claim. The core component of an argumentive essay is an evidence-based claim. Students are particularly challenged in using evidence to weaken claims, and not just to support them (Kuhn & Moore, 2015). It may be that despite their crucial importance, students are unable to envision pieces of evidence that could serve to weaken a claim. The pieces of Other-minus evidence we made available to them thus served to model this function, as well as prompt its inclusion. If students are to achieve the balanced, two-sided essays that educators want to see, they must be able to envision the evidence that would bear on the alternative they do not advocate, as well as the one they do. Once students begin to use this type of evidence in their dialogs, our results suggest, they begin to recognize its power and to use it more frequently, with it in time making its way into their writing.

In the Hemberger et al.'s (2017) study, students presented a sequence of evidence prompts showed greater use of evidence in their essays on a new topic than did students who participated in the same curriculum and also showed gains but did not receive these prompts. Here we have reported on what students receiving such scaffolding learned at a meta-level regarding understandings of evidence in argumentive writing, beyond what we have reported that they gained in performance, relative to a non-participating group. With respect to performance, in their essays the experimental group showed superior performance, relative to the comparison group, in using evidence to weaken opponents' claims, as well as in using evidence to fulfill the less challenging function of supporting their own claims. They did not, however, showed enhanced performance in the use of the most challenging evidence – evidence that supports opponents' or weakens own claims, despite the fact that examples of these evidence types had been made available to them. Such evidence, if the writer is aware of it, needs to be addressed, not ignored. Also, performance was negatively affected by a switch to a new topic where they lacked the deep engagement provided by the curriculum. There thus remains room for growth with respect to performance skills.

The present study is limited by an initial small sample size further attenuated by attrition caused by the less-than-optimal attendance typical of the inner-city public-school population studied. Confirmation of the present findings with a larger sample, as well as a more diverse one, would be desirable. Nonetheless, the posttest-only control-group design served to answer a major question the study posed. Namely, findings for the recall task rule out the interpretation that the experimental group's gains were attributable simply to the fact that they had better memory for the specific evidence available for the topic, possibly because they had engaged with the topic and were thus more familiar with it. The comparison group, who approached the topic as a new, unstudied one recalled the evidence equally well. They also showed comparable tendencies with regard to belief bias in this recall, more readily remembering evidence supportive of their position.

The results suggest, rather, that the deep engagement with successive topics, both dialogically and in writing, left the intervention group with an enhanced meta-level understanding of the role of evidence in argument. They showed greater recognition than did the comparison group of the relevance of evidence that weakened the opposing position. They did not, however, show greater skill, relative to the comparison group, in selecting specific evidence nor in anticipating the evidence they would actually use. Again, then, there remains room for growth.

Recent work by Papathomas and Kuhn (2017) shows that dialogic engagement with more capable others, as well as with peers of similar ability, enhances argumentation skill. It remains to establish what the ideal balance is between these two types of dialogic experience, but their study suggests that both have a place. It also remains at this point to identify the roles of the various components of this multidimensional and multi-phase intervention. In particular, what were the relative contributions of the

dialogic experience, leading to the "Others might say…" form that made its way into students' essays, versus the deep engagement with the topic itself. At present we see the two components as both central to the dialogic approach we have illustrated in the present work. Extemporaneous writing on a newly assigned topic is the norm in many studies of students' learning to write non-narrative essays. Outside the school context, however, they will more often have occasion to write about matters they care and have thought deeply about and likely also have engaged in discussion about with others – a further argument for studying the development of writing in such contexts.

As a bridge to individual writing, dialog has the advantage of its developmental roots in children's early conversation. The discourse skills students develop in peer dialog are of course valuable in and of themselves, not simply as a bridge to effective argumentive writing. Similarly, the knowledge students gain in arguing to learn is important in its own right (Asterhan & Schwarz, 2016), but its benefits do not end there. Arguing to learn helps develop the rigorous thinking that underlies accomplished argumentive writing. In current work, we are seeking to establish how both skill goals and knowledge goals can be pursued within the same set of activities. Both are needed if we seek to help students to produce argumentive writing that matters, to themselves and to others. None of this, of course, is to say that the only or even best path to meeting the goal of good writing is quality talk. The educational goal is such a crucial one that all potential means deserve diligent attention, in addition to the one we have described here.

Acknowledgements

We thank the principal, teachers and students of the participating school.

References

Asterhan, C., & Schwarz, B. (2016). Argumentation for learning: Well-trodden paths and unexplored territories. *Educational Psychologist, 51* (2), 164-187. https://doi.org/10.1080/00461520.2016.1155458

Barlowe, A., & Mack, H. (2002). *Looking for an argument?* New York: Teachers College Press.

Chen, Y.-C., Park, S., & Hand, B. (2016). Examining the use of talk and writing for students' development of scientific conceptual knowledge through constructing and critiquing arguments. *Cognition and Instruction, 34* (2), 100-147. https://doi.org/10.1080/07370008.2016.1145120

Cole, M. (1998). *Cultural psychology: A once and future discipline*. Cambridge MA: Harvard University Press.

Crowell, A., & Kuhn, D. (2014). Developing dialogic argumentation skills: A 3-year intervention study. *Journal of Cognition and Development, 15* (2), 363-381. https://doi.org/10.1080/15248372.2012.725187

De La Paz, S., Monte-Sano, C., & Felton, M. et al. (2017). A historical writing apprenticeship for adolescents: Integrating disciplinary learning with cognitive strategies. *Reading Research Quarterly, 52* (1), 31-52. https://doi.org/10.1002/rrq.147

Felton, M., & Herko, M. (2004). From dialogue to two-sided argument: Scaffolding adolescents' persuasive writing. *Journal of Adolescent & Adult Literacy, 47* (8), 672-683.

Ferretti, R. P., & Lewis, W. (2013). Best practices in teaching argumentative writing. In S. Graham, C. MacArthur, & J. Fitzgerald (Eds.), *Best practices in writing instruction* (2nd ed., pp. 113-140). New York, NY: Guilford.

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

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99* (3), 445-476. https://doi.org/10.1037/0022-0663.99.3.445

Hemberger, L., Kuhn, D., Matos, F., & Shi, Y. (2017). A dialogic path to evidence-based argumentive writing. *Journal of the Learning Sciences*, *26,* 575-607. https://doi.org/10.1080/10508406.2017.1336714

Kuhn, D. (2001). How do people know? *Psychological Science*, *12*, 1-8. https://doi.org/10.1111/1467-9280.00302

Kuhn, D. (2018). *Building our best future.* New York: Wessex Learning.

Kuhn, D., & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents' thinking. *Psychological Science*, *22*, 545-552. https://doi.org/10.1177/0956797611402512

Kuhn, D., Hemberger, L., & Khait, V. (2016a). *Argue with me: Argument as a path to developing students' thinking and writing* (2nd ed.). New York, NY: Routledge.

Kuhn, D., Hemberger, L., & Khait, V. (2016b). Tracing the development of argumentive writing in a discourse-rich context. *Written Communication*, 33, 92-121. https://doi.org/10.1177/0741088315617157

Kuhn, D., & Moore, W. (2015). Argument as core curriculum. *Learning: Research and Practice*, 1, 66-78.

Kuhn, D., Zillmer, N., Crowell, A., & Zavala, J. (2013). Developing norms of argumentation: Metacognitive, epistemological, and social dimensions of developing argumentive competence. *Cognition & Instruction, 31*, 456-496. https://doi.org/10.1080/07370008.2013.830618

Manz, E. & Renga, I. (2017). Understanding how teachers guide evidence construction conversations. *Science Education, 101* (4), 584-615. https://doi.org/10.1002/sce.21282

Mcneill, K., & Berland, L. (2016). What is (or should be) scientific evidence use in K-12 classrooms? *Journal of Research in Science Teaching, 54* (5), 672-689. https://doi.org/10.1002/tea.21381

Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach.* New York: Routledge. https://doi.org/10.4324/9780203946657

Newell, G.E., Beach, R., Smith, J., & VanDerHeide, J., (2011). Teaching and learning argumentative reading and writing: A review of research. *Reading Research Quarterly, 46* (3), 273-304.

Nussbaum, E.M. (2008). Using argumentation vee diagrams (AVDs) for promoting argument-counterargument integration in reflective writing. *Journal of Educational Psychology, 100* (3), 549-565. https://doi.org/10.1037/0022-0663.100.3.549

Nussbaum, E. M., & Asterhan, C. S. C. (2016). The psychology of far transfer from classroom argumentation. In L. Resnick, C. Asterhan and S. Clarke (Eds.), *The psychology of argument: Cognitive approaches to argumentation and persuasion* (pp. 407-423). London: College Publications.

Nussbaum, E.M., & Edwards, O. (2011). Critical questions and argument stratagems: A framework for enhancing and analyzing students' reasoning practices. *Journal of the Learning Sciences, 20* (3), 443-488. https://doi.org/10.1080/10508406.2011.564567

Olson, D., & Oatley, K. (2014). The quotation theory of writing. *Written Communication, 31* (1), 4-26. https://doi.org/10.1177/0741088313515164

Papathomas, L., & Kuhn, D. (2017). Learning to argue via apprenticeship. *Journal of Experimental Child Psychology, 159,* 129-139. https://doi.org/10.1016/j.jecp.2017.01.013

Resnick, L. B., Michaels, S., & O'Connor, C. (2010). How (well structured) talk builds the mind. In R. Sternberg & D. Preiss (Eds.), *From genes to context: New discoveries about learning from educational research and their applications.* New York: Springer.

Resnick, L., Asterhan, C., & Clarke, S. (Eds.). (2015). *Socializing intelligence through academic talk and dialogue*. Washington DC: American Educational Research Association. https://doi.org/10.3102/978-0-935302-43-1

Reznitskaya, A., Anderson, R., McNurlen, B., Nguyen-Jahiel, K., Archodidou, A., & Kim, S. (2001). Influence of oral discussion on written argument. *Discourse Processes, 32* (2-3), 155-175. https://doi.org/10.1207/S15326950DP3202&3_04

Reznitskaya, A., & Wilkinson, I. (2017). *The most reasonable answer: Helping students build better arguments together.* Cambridge MA: Harvard Education Press.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge MA: Harvard University Press.

van Eemeren, F., & Grootendorst, R. (1992). *Argumentation, communication and fallacies*. Hillsdale NJ: Erlbaum.

Vygotsky, L. S. (1978). *Mind in society: the development of higher psychological processes.* M. Cole, V. John-Steiner, S., Scribner, E. Souberman (Eds). Oxford, England: Harvard University Press.

Walton, D. (2014). *Dialog theory for critical argumentation.* Amsterdam: John Benjamins.

Zillmer, N., & Kuhn, D. (2018). Do similar-ability peers regulate one another in a collaborative discourse activity? *Cognitive Development, 45*, 68–76. https://doi.org/10.1016/j.cogdev.2017.12.002

Appendix A: Questions and Answers about Juvenile and Adult Court

1. Q: What are public opinions on the juvenile court issue? (A+)
A: People hold different opinions on this issue. However, a "get tough" policy has become more popular in recent decades, with almost every state passing laws in the 1990s making it easier to try juveniles in adult courts.
2. Q: At what age is the brain fully developed? (J+)
A: The prefrontal cortex, which is responsible for abstract thinking and the ability to exercise good judgment, is not fully developed until about the age of 25.
3.Q: Do adult jails provide job training? (A+)
A: Yes, most adult jails teach job skills to help prisoners earn a living when they are released.
4. **Q: Can teens continue their education while at a Juvenile Detention Center? (J+)**
A: Juvenile centers provide some schooling, but it may not be a full day or every day. But teens are likely to get better general education at a juvenile center than an adult prison.
5. Q: Are teens at risk of being assaulted in adult prisons? (A-)
A: Yes. Teens in adult jails are 50% more likely to be attacked by another inmate and twice as likely by prison staff, compared to adult prisoners.
6.Q: Do all courts give the right to a trial by jury? (J-)
A: No. Juvenile courts don't allow trial by jury. A judge hears evidence and rules.
7. Q: How many murders are committed by teens? (J-)
A: In 2008, 9% of murders in the US were committed by juveniles.
8. Q: Do prisoners have counselors to talk to?
A: They may have a counselor to talk to. However, this is more common in juvenile than adult prison.
9. Q: Are teens likely to repeat their crimes?
A: For teens convicted of a felony, the rate of recidivism (repeat crime) is 90% over 10 years. For crimes overall, it is about 50%.
10. Q: Are the sentences given for crimes less harsh in juvenile than adult court?
A: Compared to adult court sentences, juvenile court sentences tend to be less harsh, with probation and parole more likely.
11. Q: What proportion of violent crimes are committed by juveniles?
A: Juveniles were involved in one-quarter of violent crimes over the last 25 years.
12. Q: Do teens that go to jail get jail records?
A: They do not if sentences are served in a juvenile detention center; their records are sealed on release.