Apprenticeship As a Developmental Mechanism in Argumentation Skill Development

Yu Song

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2024

Abstract

Apprenticeship As a Developmental Mechanism in Argumentation Skill Development

Yu Song

Argumentation is widely regarded as both a productive path and a critical objective of education. However, poor performance remains a serious problem at all ages in assessments of expository writing in which students are asked to make an argument in support of a claim. An apprenticeship model is proposed as a mechanism in the development of skill in dialogic argumentation, with this skill serving as a bridge to individual written argument. In a multi-week intervention, young adolescents were paired with a series of both more skilled and similarly skilled partners, anonymously, in conducting one-on-one electronic dialogs on controversial issues. A comparison group was included who engaged in the same intervention and assessments, but their dialogic partners were confined to similar ability peers. The more skilled adult partner displayed skilled forms of counterargument and use of evidence to support claims as well as frequent questioning with respect to the partner's statements and meta-talk about the discourse itself. Effects on students' individual argument skill on a new topic were assessed by means of both a (solitary) individually constructed dialog and an individual essay. In both the dialogs and essays, the experimental group showed greater skills in using evidence to support a claim, generating advanced counterarguments, and constructing integrative critical arguments coordinating two contrasting claims, relative to the comparison group. These results lend support to the power of apprenticeship in individual argument skill development. Both groups also advanced in individual dialogic argument skills following their engagement in argumentation, a

result thereby demonstrating the passage of higher-order intellectual skills from a social to individual level. Besides their educational implications, the theoretical significance of these results in relation to both an apprenticeship model and a dialogical model of argument skill development was discussed.

Table of Contents

List of Tables

List of Figures

Acknowledgments

The journey of PhD study is not an accomplishment of myself. I can never achieve this dream without the help, support, and encouragement I received from endless individuals in the past three and a half years. Upon the completion of my degree, I want to take the time to express my gratitude to them.

First and foremost, I want to express my deepest appreciation to my advisor, Dr Deanna Kuhn. Professor Kuhn, you saw my potential to grow and took me as your doctoral student three years ago. You gave me the freedom and flexibility to explore my research interest and guided and supported me through every step of my PhD study, even as our lives were disrupted by the Covid-19 pandemic. Your unparallel professionalism and expertise in argumentation and scientific thinking have not only strengthened my understanding of these crucial thinking skills but also prompted me to reflect on and develop my own thinking and reasoning practices. I admire your insightful take on contemporary educational issues, keen identification of major research questions/findings, and unwavering dedication to details, precision, and concision. I aspire to move on with my own work and life with the attitudes and passion you have shown me.

I am also deeply grateful to my thesis committee members. Dr. John Black, Dr. Tyler Watts, Dr. Bryan Keller, and Dr. Julia Zavala, you kindly agreed and made time to participate as my thesis committee members. Throughout the process, your perceptible and stimulating questions and comments on my thesis project pushed me to refine and enrich it and to reflect on the practical and theoretical implications of my findings on broader and cross-cultural levels. I would not be able to complete my dissertation project without your generous help and support.

I also want to extend my genuine gratitude to Dr. Yuchen Shi. You are like my big sister who introduced me into the field of argumentation research. Working side-by-side with you at

ECNU in the summer of 2019 and later on other projects have revealed to me your passion for and commitment to advance educational research in China. Your enthusiasm, curiosity, and hardworking have motivated me to explore my own passion in work and life.

I must also thank the students and teachers from the school where I conducted my intervention study. Principle Wang, Teacher Shi, Teacher Juan, Teacher Hui, and Teacher He, you gave me incredible trust and assisted and supported me whenever I was in need. Teacher Li, I could not complete this project without your help. I admire and aspire to apply the same level of organizational skills and conscientiousness you shown me when we worked together. The kids who participated in my study, you welcomed me with the warmest smile, engaged in each phase of my program with great attention and curiosity, and took each assessment task seriously. Thank you all for making my dream come true.

I also feel grateful to meet and study with three other fellow students of Dr. Deanna Kuhn. Si Xiao, we started the program together and became good friends while we both navigated our studies and lives under the shadow of the Covid-19 pandemic. We share so many feelings and I feel incredibly lucky to have earned such a good friend on this journey. Teresa Fraguada and Sybille Bruun, we went through the same procedures of PhD training and shared encouragement and information throughout the process. I am impressed by your passion for research and education. Thank you for keeping me accompanied and motivated along the way.

The list certainly does not end here. For the endless individuals who ever pat my back or cheered me up, your kindness is invaluable to me in the past and for the future.

Yu Song

On Nov 23, 2023

Dedication

This thesis is dedicated to my family.

To my parents, Xuefeng Li and Jiangxiao Song, for your unconditional love and support, your unparallel compassion and patience, and your demonstration of entrepreneurship and courage in your own lives.

To my husband, Shawn Ling, for your always prioritizing my happiness, responding to my needs with love and empathy, and becoming my best friend and lifemate.

To my parents-in-law, Liya Zhang and Jianping Ling, for your love and care for me as your daughter and your dedication to exemplary moral and professional standards.

I love you all, with all of my heart.

Introduction

The significance and challenges of developing Argumentive[1] competence

Today the rapid advancement and expansion of knowledge supported by breathtaking technological evolutions have complicated the challenges faced by educational researchers, practitioners, and policy makers alike as they grapple to understand how to prepare students for the future, which is identified as the key mission of education in the 21-century world (OECD, 2020). With the task of predicting which portions of today' knowledge will still be relevant in the future and thus worth imparting to students deemed impossible to accomplish, the significance of educating for thinking has gathered increasing steam. Skill in thinking enables individuals who master it to learn and update their knowledge and understanding in response to the everchanging requirements of the environment. It is in this sense that education for thinking has begun to appear a more sustainable goal.

The next challenge is to define and describe the characteristics of skilled thinking and how to identify, assess, and develop it in schools, the most promising and appropriate institutions to nurture the minds of our younger generation. The perspective adopted here is that skilled thinking can be defined as thinking aimed at advancing knowledge by means of consciously controlled coordination of theory and evidence. As such, it is closely aligned to scientific thinking, which shares these aims (Kuhn, 2010). Scientific thinking and practice include two major dimensions, inquiry and argument. Inquiry encompasses purposeful activities to gather, compare, interpret, and evaluate relevant data that are initiated by and ultimately answer a

[1] Argumentive was coined by Deanna Kuhn in her pioneering book *The Skills of Argument* and used by many others in related fields since then to stress students' abilities and dispositions to develop/elaborate arguments within argumentative contexts.

meaningful question. The second dimension, argument, introduces resulting claims to a community of others to establish their merit within a framework of evidence and alternatives. Argumentation typically builds upon and extends the achievement of inquiry learning by seeking to coordinate new knowledge with prior beliefs and/or alternative explanations to justify a conclusion. Thus, inquiry can be regarded as the input component while argumentation the output component of knowledge-seeking activity. Inquiry thus provides the groundwork for argumentation in terms of both content and skills (Kuhn, 2001).

Teaching and learning as argument are now widely advocated by theorists and researchers in educational and cognitive psychology as a promising pathway for expanding students' knowledge base as well as developing their minds (Andriessen et al., 2003; Howe & Zachariou, 2019; Iordanou et al., 2019; Nussbaum, 2021; Kuhn, 2019; Kuhn & Moore, 2015). This attention was echoed by the Next Generation Science Standards (NGSS Lead States, 2013) and the Common Core (CCSSI, 2010) in emphasizing argumentive competence as a key objective of education. Contemporary research efforts to promote argumentive skills in educational settings can be categorized into two parallel lines of work –*argue to learn* and *learn to argue*. The former emphasizes argumentation as a tool to promote deep understanding of a concept and facilitate conceptual change, with gains in the skill of argument a derivative of content learning. The latter, in contrast, identifies argumentation as a sufficiently critical and complex cognitive and social practice to deserve dedicated and extended educational investment of its own. The present work fits with the *learn to argue* perspective and seeks to advance understanding of the mechanisms of the development of argumentive competence with the goal of better supporting adolescents' mastery of it.

As people engage in argumentation (either collaboratively with real others or individually with a generalized hypothetical other), two weaknesses have been repeatedly observed from childhood through adolescence and adulthood, disconfirming the assumption that argumentive competence is innate or develops naturally with age. One concerns the dialogic nature of argumentation and manifest itself as exclusive concentration on exposition of one's own ideas to the exclusion of attending to and addressing the opposing side (Felton & Kuhn, 2001). Failing to engage with the opponents' arguments violates the two goals of argumentation identified by Walton (1989) – one is to secure commitment from the opponent that can be used to support one's own argument and the other is to weaken the argument of the opponent by identifying and challenging its weaknesses. Both emphasize the necessity of addressing the opposing position in argumentation.

The second weakness hinges on the epistemic status of an argument as a claim justified by relevant evidence. People typically struggle to base their arguments on sufficient and relevant evidence and/or to provide proper warrants to establish a relation between argument and evidence (Macagno, 2016). However, coordinating arguments with evidence bearing on them is central to establishing the tenability of an argument and thus key to productive and constructive argumentation (Macagno, 2016; Kuhn, 2001; Zimmerman, 2007).

Mechanisms in the development of argumentive competence

An increasing number of researchers in the fields of educational, developmental, and cognitive psychology have undertaken to explore and assess approaches to promote the development of argumentive competencies in adolescents (Nussbaum, 2008, 2021; Sandoval & Millwood, 2005; Resnick et al., 2018; Reznitskaya & Gregory, 2013). Among them, the dialogic approach, developed and further refined by Kuhn and colleagues (Crowell & Kuhn, 2014;

Hemberger et al., 2017; Kuhn & Crowell, 2011; Kuhn et al., 2016; Shi, 2019), has received considerable attention among researchers in both Western and Eastern cultures. In general, researchers who have implemented this approach in both cultures report advancements among participants following extended interventions. These include not only generation of more advanced counterarguments (Kuhn & Crowell, 2011; Kuhn et al., 2016), but also coordination of arguments with evidence of varying functions and forms to strengthen their arguments and weaken those of opponents (Hemberger et al., 2017; Kuhn & Moore, 2015; Shi, 2019; Shi et al., 2019).

Regarding the mechanisms of development of argument skills, an activity that has been shown to support such development is participants' extended and dense engagement in direct argumentation with peers (Kuhn, 2019). This finding is in line with the sociocultural perspective that emphasizes learning as apprenticeship into communities of practice (Reznitskaya, et al., 2007; Rogoff, 1990; Sandoval & Millwood, 2005; Vygotsky, 1978). In other words, learning is perceived as a process of social regulation that becomes interiorized on an individual plane. Acquiring argumentation through cognitive apprenticeship encompasses the means and standards of defending and challenging a position.

This perspective has motivated an emerging line of research on argumentation skill that focuses on apprenticeship as a mechanism of acquisition of argumentation skill (Halpern, 2022; Mayweg-Paus et al., 2015; Papathomas & Kuhn, 2017). These experimental studies have shown that gains in argumentation skill are greater when novices are paired with more expert arguers rather than only with peers. It is argued that the more capable interlocutor provides support for the less capable one, creating a *zone of proximal development* (i.e., the distance between the less

capable interlocutor's actual ability and potentially achievable ability modelled by the more capable interlocutor).

The dialogic argumentation approach draws further support from a Piagetian sociocognitive perspective which emphasizes development as a process of *self-regulation*, albeit influenced by social and cultural factors. The extended and dense engagement in authentic peer-to-peer argumentative dialogues afforded by the dialogic argumentation approach provides students with plentiful opportunities to reflect on and coordinate their reasoning as they debate both agreeing and disagreeing peers, in so doing facilitating development (Crowell & Kuhn, 2014; Iordanou & Rapanta, 2021). This account inspired a related line of research focusing on promoting students' meta-level awareness of the nature and function of argument, as a means to enhance their argumentative performance. Exemplifying the varying functions of evidence in relation to claims promoted students' attention to and addressing of other-side evidence in their argumentative essays (Hemberger et al., 2017). Additional prompt to engage in meta-level reflection on the nature of evidence and its relations to various claims produced gains in students' ability to address evidence inconsistent with their chosen position (Shi, 2019; Iordanou, 2022).

The present study and research questions

Experimental studies cited above have established that dense engagement and practice with peers advances development of argumentive skill. However, Children and adolescents engage in argumentive discourse in their daily lives not only with peers but also with older individuals possessing more advanced skills. How in the latter case do the young who are exposed to these displays of more advanced strategies attend to and interpret them, eventually incorporating them into their own repertory to see how they fare? As an attempt to seek answers

to these questions, the present study was so designed that electronic discourse takes place both between two peer-pairs but also between a peer-pair and an older, more skilled discourse partner. The electronic medium allows the substitution of a more skilled dialogic partner to occur without the awareness of the target participants, excluding the confound of social factors. The design also includes a comparison condition whose participants engage in dialogs only with similar-ability peers.

The hypothesis is that both groups will show increasing argumentation skill over time, but this progress will be greater in the condition that includes a mixture of more and less skilled partners. This hypothesis has received some support in previous studies. One of them involved comparisons only at a local episodic level (Mayweg-Paus et al., 2015) in the absence of longer-term engagement. The two others showed enhanced argumentation skill over time among the experimental group, relative to a peer-only comparison group, over a period of one to several months (Halpern, 2022; Papathomas & Kuhn, 2017). But none of them took the important step undertaken by the present study to show that skill advancement occurs in individual written argument, besides dialogic argumentation. It is this demonstration that will show that skills have truly passed from the social to the individual plane, a demonstration that carries critical weight to support both a dialogic and an apprenticeship mechanisms of argumentation skill development.

Chapter 1: Literature Review

1.1 Dialog as Path to Individual Written Argument

Students of all levels from early elementary through post-secondary perform poorly on assessments of expository writing in which they are asked to make an argument in support of a claim (Graham & Rijlaarsdam, 2016). This prompted many researchers to investigate approaches to improve students' argumentive skills, given their significance for the success of an individual and the development of a healthy democratic society. However, many of these attempts only received at best modest results with limited generalizability. Thus, before succumbing to any one of them, it is worth a concerted effort to learn all possible about the developmental mechanisms that enable novice arguers to progress to skilled ones, as a foundation for determining how best to support this development. The work presented here has this objective.

Dialogic argumentation has been proposed (Kuhn et al., 2016) as a fruitful path to individual written argument. Rooted in the everyday talks people engage in with others, dialogs have the potential to evolve into a more formal, symbolic, and intrapersonal form manifested as elaborated arguments generated by individuals in relation to specific argumentative purposes (Kuhn & Crowell, 2011; Kuhn et al., 2016). However, dialogic argumentation has received less attention than argumentive writing, at least in the precollege curriculum, even though the close connection between the two forms of argument has ancient roots and is now well noted (Billig, 1987; Gergen, 1992; Kuhn, 1991; Mercer, 1995).

A dialogic approach to thinking and writing can be traced back as far as to Socrates and Plato in their pursuit of the ultimate truth through critical argumentation and Aristotle's development of logic, dialectic, and rhetoric. In contemporary philosophy, Mead was seen as a

significant figure who rejuvenated people's interest in dialogic method by describing thought as "the conversation of the generalized other with the self" (Mead, 1922/1934, p264); thus, thinking is seen as dialogic in nature. Further development can be attributed to Grice (1975), Van Eemeren and colleague (van Eemeren & Grootendorst, 1992), and more recently, Walton (1989, 2014), who underscored that the analyses and assessments of arguments should be situated within specific dialogic contexts. Most recently, there has even emerged a trend in uniting all sorts of thinking under the umbrella concept of argumentation that arguably captures the evolutionary functions and purposes of thinking (Kuhn, 2019; Mercier & Sperber, 2011).

Within psychology, Vygotsky's sociocultural theory laid the groundwork for the conception and later proliferation of dialogic teaching/learning methods. According to him and his successors (Rogoff, 1990; Vygotsky, 1978), all higher-order cognitions (i.e., cognitions mediated by psychological tools like language) originate in the social contexts and the inter-mental processes with practice will interiorize and be appropriated by the individual. Though not his major concern, Piaget also acknowledged the values of social intellectual exchanges in initiating and promoting individual cognitive development. The major factor he pinpointed to that can account for these effects was the cognitive conflicts arising from discussions. Cognitive conflict prompts the discourse partners to work inwardly to resolve disparities and restore individual equilibration at a higher-level.

Based on empirical observations and investigations, Graff (2003) claimed that the strength of a dialogic approach to writing resided in its ability to provide the otherwise "missing interlocutor". In the absence of an interlocutor, students' argumentative reasoning is always afflicted with biases and laziness (Mercier et al., 2016). When asked to express themselves alone in writing, they tended to focus exclusively on elaborating their own positions and/or subject the

opposing position to stricter evaluative standards (Hemberger et al, 2017; Shi, 2019). The benefits of interlocutors lie in that they serve as audience to which the writing is addressed. In other words, the presence of an interlocutor provides the writer with a purpose to consider and address alternative points (Kuhn & Moore, 2015).

On the cognitive fronts, researchers have proposed that the benefits of dialogic engagement to individual writing performance lie in students' acquisition and consolidation of a series of transferrable cognitive skills implicated in both contexts. Among others, proactive executive control strategies are identified as the ones doing the heavy lifting (Nussbaum & Asterhan, 2016). During argumentation, students practice and strengthen executive control strategies that are intentionally activated to serve specific argumentative purposes (e.g., attacking the opposing side by implementing a complex argument scheme like cost-benefit analysis). These are proactive executive control strategies (PECS), which contrast with reactive strategies that are activated only after the presence of a particular stimuli. In the process of argumentation, participants need, at a minimum, to switch back and forth between evaluating, critiquing, and defending alternative claims while inhibiting any interference from and maintaining the goal of their own argument. This provides them with plentiful opportunities to practice and gain facility in implementing various PECS. As a result, the cognitive costs of implementing them reduce and the possibilities of them being activated in writing tasks increase through spreading activations.

A sizable body of empirical studies have demonstrated that rich experience in one-on-one dialogic argumentation serves as a bridge to individual argumentative writing (Hemberger et al., 2017; Iordanou et al., 2019; Kuhn, 2019; Kuhn & Crowell, 2011; Kuhn et al., 2022; Matos, 2021; Rapanta & Felton, 2021; Shi, 2019; Shi et al., 2019). Specifically, researchers have found that argumentive skills tended to first appear in dialogs and then gradually transmit to individual

written arguments (Hemberger et al., 2017; Shi et al., 2019). Following their argumentation practices, features of dialogic exchanges appeared in students' writings. For example, students introduced alternative views by stating that "the opponent might say…" (Kuhn et al., 2016). Additionally, collaborative writing between opposing side students was found to be a facilitator of individual integrative writing that leveraged on the presence of an explicit (opposing) interlocutor (Matos, 2021).

The close connection between dialogic and individual written argument enhances the educational significance of skill in discourse, to the extent that it is both a fruitful path to writing and a meaningful objective in itself. We know that argumentive discourse undergoes enormous development, going back as far as a young child's verbal objection to a parent's directive and from there to a young adult's well-argued case establishing why their chosen course of action is preferable to an alternative. However, the mechanisms underlying this notable course of development are not well understood which limited its educational implications. In the present study, we consider the mechanisms that enable the development of this critical skill, one especially notable given that it is normally not explicitly taught, unlike expository writing.

In the subsequent sections, we will first clarify the conceptual meaning and empirical operationalization of argument to lay the groundwork for the present study. Next, we will proceed to review and discuss pedagogical approaches to foster argumentive skills with a particular focus on the "Argue with Me" (AWM) approach developed and popularized by Kuhn and Colleagues and implemented in the present study. Finally, after foregrounding the mechanism issue, we will conclude this chapter by drawing out our research questions and hypotheses.

1.2 The Assessment of Argumentive Competence

Researchers have been using the word argument to refer to two meanings that are distinguishable but not always explicitly differentiated in practice. One the one hand, argument can be understood as a rhetoric product consisting of, minimally, a proposition and, preferably, supporting reasons and/or evidence bearing on it. On the other hand, argument can also refer to a dialectic process people engage in with others that is the practice of argumentation (Macagno, 2016; Rapanta & Felton, 2021).

Correspondingly, argumentive competence can be conceptualized onto two dimensions: one is the construction of structurally sound arguments (argument-as-a-product view) and the other one is individuals' relevant participation in argumentation (argument-as-a-process view) (Macagno, 2016). Educational researchers aiming to foster students' argumentive skills typically focused on either one of them with the argument-as-a-product view receiving lion's share of attention (see Rapanta & Felton, 2021, for a review). Along this line of work, researchers typically base their characterizations and evaluations of argument quality on Toulmin's Argument Pattern model (TAP, Toulmin, 1958/2003). According to Toulmin, an argument can be represented by an interconnected set of elements. Three of them are most essential: a claim, data, and a warrant linking the claim and data. Other components include a backing to substantiate the warrant, a qualifier to set limits for the argument, and a rebuttal to address alternatives. On this perspective, the quality of an argument is determined by its structural complexity; in other words, the more components an argument contains, the stronger it is. Researchers within this camp typically strive to scaffold student's construction of complex, multi-component arguments by providing them with various visual/graphic organizers listing and

linking possible elements of an argument (Muller-Mirza et al., 2007; Nussbaum, 2008; Ryu & Sandoval, 2012).

Alternatively, students' argumentive competence can also be assessed relative to the expected mode of participation (Macagno, 2016). Argumentation can occur on the social as well as the individual planes, but regardless of the contexts, dialogicality (Koshmann, 1999) is an inherent feature of a sound argument. Dialogicality implicates intersubjectivity and alterity. The former requires establishment of shared understandings among participants and the latter requires bracketing and distinguishing one's own ideas from those of others. On the process-based perspective, an argument should always be locally relevant and globally informative in relation to the argumentative goals it serves. The essential function of an argument, thus, can be characterized as to modify (either to increase or decrease) the acceptability of a perspective within the context of argumentation (Macagno, 2016).

Following this perspective, to assess argumentive skills within a social dialogic context, whether a new contribution by a participant is embedded within the ongoing dialog by either expanding on or countering what have already been said is a key question. Counterargument is a critical strategy in argumentative dialogs as it not only responds to the interlocutor's preceding argument relevantly but also advances the epistemic goal of the dialog (i.e., advanced joint understanding). This can be accomplished by either a single move that directly challenges the conclusion or premise of the opponent's preceding argument or several intentionally planned consecutive moves to first uncover and then attack the interlocutor's implicit commitments like questioning or meta-talk statements followed by an undercutting move (Kuhn & Felton, 2001).

On the individual plane, expository writing is a classic means to externalize intra-mental argumentation (or argumentative reasoning). An argument can be locally relevant as long as its

content is connected and coherent to the content of adjacent arguments advanced by the speaker. The occurrence of an unconnected unit within an argumentative text is an example of an ineffective move which interrupts the dialogic flow of reasoning. The relevance of an argument also hinges on its pragmatic function in relation to the central claim the interlocutor seeks to advance (i.e., how informative an argument is to support that claim). An argument can take on various forms to accomplish the same goal. For example, to increase the strength of a favored position, an argument can be constructed to either support that position or weaken an opposing position.

Beyond strategic level performance, students' metacognitive competence and dispositions including their meta-strategic knowledge of the purposes, norms, and various strategies of arguments as well as epistemological understandings of the nature of knowledge and the process of knowing as argumentation are critical aspects of argumentive competence. They regulate and constrain students' dialogic behaviors to achieve productive and constructive argumentation (Iordanou & Rapanta, 2021; Kuhn & Zillmer, 2015; Rapanta & Felton, 2021).

1.3 Pedagogical Approaches to Learning to Argue

With the recognition that argumentive competence are transferable to impact a wide range of learning, logic/reasoning, and general intelligence performance (Asterhan & Schwarz, 2016; Resnick et al., 2018), there has been a growing line of work devoted to developing students' argumentive skills in schools (see Rapatan & Felton, 2021 for a review). The purpose of this work can be summarized as "learning to argue" (LTA), which distinguishes itself from another related area of research that is referred to as "arguing to learn" (ATL). The latter also

stresses the importance of argumentation in classrooms, but it concentrates on utilizing argumentation as a pedagogical tool in service of other educational purposes like the learning and understanding of subject matters (see Asterhan & Schwarz, 2016 for a review). By contrast, LTA research views argumentation as a core competence and educational objective on its own.

In the present study, we align ourselves with the LTA perspective because we believe that given the complexity and significance of argumentive competence, it requires and deserves dedicated efforts to be nurtured within supportive environments. As discussed above, building on the theoretical legacy of Vygotsky's (1978) sociocultural perspective, which emphasizes the functions of language as both a cultural tool (for constructing and sharing knowledge) and a psychological tool (for organizing and structuring thinking and reasoning), as well as the close connection between the social and the psychological (Billig, 1987), dialog has been foregrounded as a promising starting point and pathway for developing individual argumentive competence (Resnick et al., 2015, 2018). It's believed that through language-mediated joint reasoning activities, the "inter-mental" processes with practice become internalized to shape some aspects of "intra-mental" reasoning (Kuhn & Crowell, 2011, Mercer et al., 2017; Shvarts & Bakker, 2019). For this socially rooted learning to occur, however, extended involvement in dialogs with others is indispensable. Repeated practices not only provide opportunities for participants to strengthen new skills but also are required for them to discover and become committed to the values of argumentation so that they are disposed to exercise it when instructional prompts/cues are withdrawn (Kuhn, 2005).

Dialogs can take many forms. Which kind of dialog has the greatest educational potential? Researchers diverge in their conceptualizations of effective/desirable educational dialogs, which result in a mixture of intervention programs with idiosyncratic contents,

structures, and outcomes. To promote an integrated and coherent understanding of the efficacies of the dialogic method in developing students' argumentive competence, Rapanta and Felton (2021) reviewed and analyzed the methods (i.e., the framing and fostering method of effective dialogs), dialog goals (i.e., the epistemic goal that participants strive to achieve though dialogs), LTA outcomes (including outcomes related to both argument as a product and as a process perspectives), as well as the learner and learning contexts of 143 empirical studies. They found a significant relationship between the framing methods and the dialog goals across different studies and derived four major framing patterns that are representative of the wide range of methodologies developed by different researchers. The four patterns – low-structured sense-making, high-structured articulation, high-structured deliberation, and dialectic deliberation dialogs – have profound impacts for how the learning environments are structured and what LTA outcomes are expected/achieved. The three framing approaches, low-structured, high-structured, and dialectic, differ on the extent to which dialogic activities are intentionally orchestrated to elicit (adversarial) argumentation among students to induce argumentive gains. The three epistemic goals – sense-making, articulation, deliberation – represents how deeply alternative opinions or theories are contrasted, discussed, and coordinated.

Low-structured sense-making dialogic method stresses the formative potentials of educational/classroom dialogs. It is low-structured because it does not prescribe any argumentative activities, rather students are encouraged to participate in discussions freely and actively. They are expected to learn about, construct, and accommodate their behaviors to the norms of productive dialogic interactions through active participation. For this learning to happen, however, teachers play significant roles. They have to mediate student dialogs by promoting effective dialogic moves while discouraging less effective or desirable ones. In the

process of classroom discussions, students co-construct meaning and share interpretive authority among peers and teachers alike. The "exploratory dialog" (e.g., Mercer, 2004) is an example of such dialogs. Through participation, LTA gains are believed to emerge naturally as students acquire the norms of effective dialogs (Mercer et al., 2017) and display more advanced modes of participation like elaborated reasoning (Reznitskaya et al., 2012), divergent thinking (Damico & Rosaen, 2009), and consideration of alternative perspectives (Mercer et al., 2017). However, evidence regarding whether characteristics of dialogic discussion can be transferred to impact individual reasoning practices is sparse and appeared to be inconsistent. Reznitskaya and colleagues (2012) implemented the Philosophy for Children (P4C) program that successfully induced productive inquiry dialogs among students. Over time, students were observed to exercise more control over their dialogic exchanges and display more advanced dialogic moves such as elaborated explanations and reasoning in dialogs. However, these improvement in dialogs did not get transferred to individual argumentative reasoning tasks when they were asked to express themselves alone. The effects of whole-class low-structured discussions appear to be unstable and narrow, and this is partly because of the uncertainties related to the participation and thus gains of each individual student especially those who tend to fade into background.

High-structured dialogic method, by contrast, seeks to pursue explicit learning goals regarding the knowledge and skills of argumentation by engaging students in well-structured activities and tasks intentionally designed to elicit meaningful argumentative interactions and outcomes (e.g., Ford, 2012; Golanics & Nussbaum, 2008; Hsu et al., 2015; Sampson & Clark, 2011). The learning conversations are framed in relation to various epistemic goals like articulation and deliberation. Regardless, students are typically organized into small groups so that they have more spaces and opportunities to talk. Though studies adopting the high-structured

dialogic method often observed transfer of skills from the dialogic to the individual reasoning contexts (e.g., Cross et al., 2008), they preferred the TAP model to gauge LTA gains and had limited attention on the relevancy aspect of arguments.

The dialectical deliberation approach is a special case of the high-structured method in the sense that the learning dialogs are highly choreographed to induce adversarial argumentation as a particular type of dialogic exchanges that is believed to be most beneficial for developing participants' argumentive competence. Such dialogs usually take place in one-on-one (either between students or dyads of students) settings and with the support of computers to create an optimal collaborative learning environment. Students come into the conversations with different perspectives that are made clear to each other at the beginning and seek to reach an agreement at the end. As they are confronted with, deliberate, critique, and ultimately seeks to reconcile contrasting arguments and evidence to arrive at a single "best explanation", they acquire, practice, and strengthen various argumentive strategies (e.g., counterargument and rebuttal) as well as metacognitive competence and dispositions (Kuhn & Zillmer, 2015; Kuhn, 2022). The dialectic nature of adversarial argumentation renders it the most challenging one for students to engage in and master. This is because in order to successfully defend their own position, they have to constantly switch perspectives and apply execute controls (e.g., inhibiting the interference of and bracketing alternative arguments) to consider and reason about both their own and the opponent's arguments within a framework of alternatives and evidence (Kuhn & Crowell, 2011). But as they gain facility in participating in such highly demanding dialogs, they also reap the most benefits from it at the cognitive, metacognitive, and epistemological levels (Iordanou & Rapanta, 2021; Kuhn, 2015; Rapanta & Felton, 2021). Transfer of gains from

dialogs to individual reasoning tasks have been consistently documented, though with varying extents depending on the levels of scaffolding and/or amount of exercises.

1.4 "Argue With Me": An Argumentation-based Curriculum

The AWM approach we employ in this study fits best with the dialectic deliberation category. It features student dyad's dense and extended engagement in (adversarial) argumentative dialogs directly with an opposing-side dyad. Dyadic argumentation supplemented by other collaborative/reflective activities has been proven to induce and develop a host of argumentive competencies of students (see Iordanou and Rapanta, 2021 for a review). Microgenetic studies have been implemented to trace the developmental trajectories of new skills. The general idea behind microgenetic studies of cognitive skill development is that dense engagement and practice over an extended period will accelerate the development of an otherwise naturally occurring process, allowing the researcher to gain better understanding of the process (Kuhn, 1995; Siegler, 2006). With extended practice, it's been observed that students gradually discover and apply more advanced strategies more frequently and consistently (Meyweg-Paus et al., 2015; Papthomas & Kuhn, 2017). These stronger strategies gradually come to dominate weaker ones, a process implicating advancement at and support from the metacognitive as well as epistemological levels.

Having been implemented and empirically tested in over 30 studies involving adolescents from the United States, Europe, and Asia, the AWM approach has received consistent and remarkable success in developing students' argumentive skills through engaging them in dialogic argumentation (see Iordanou & Rapanta, 2021 for a review). Among others, three aspects of

development have been consistently documented that are considered central to productive and constructive argumentative reasoning.

One aspect of development concerns with students' ability to coordinate arguments with evidence bearing on them (Hemberger et al., 2017; Shi, 2019; Shi, et al., 2019). Kuhn and colleagues' conceptualization of evidence-based argument goes beyond Toulmin's structural model in that they emphasized the relevance of evidence use. Relevant or functional use of evidence is differentiated from non-functional use of evidence. To make such a differentiation, they first categorized information and evidence as different epistemic entities. Specifically, information (i.e., raw data or factual knowledge) carries no meaning in itself and only becomes evidence when it is explicitly linked to and successfully functioned to support a claim. Here the emphasis is on the warrant and/or backing connecting the claim with the information as evidence, beyond and above the quality or source of the information. As a result, the sources from which children can draw evidence to support their arguments (or in other words, to construct functional evidence-based argument) expand to include their personal knowledge base. This enabled researchers to trace children's developing epistemological theory of evidence (Fedyke, et al., 2019), taking into account how students' pre-argumentation forms of epistemology gain opportunity to emerge and develop through school-based activities (Iordanou & Rapanta, 2021).

Studies employing the AWM approach have found that students improved their ability to coordinate arguments with relevant evidence following dialogic argumentation practices (Kuhn & Moore, 2015). Regarding the trajectory of development, evidence was found to be more quickly and frequently employed to support students' own or weaken the opponent's position while more gradually and to a lesser extent used to acknowledge the merits of the opposing

position or the weaknesses of their own position (Hemberger et al., 2017; Shi, 2019; Shi, et al., 2019). Accompanying this development are students' increased appreciations of the values and understandings of the functions of evidence (Hemberger et al., 2017; Kuhn & Moore, 2015; Shi, et al., 2019) and their more frequent and reciprocal meta-talk regarding the norms and standards of evidence-based arguments (Kuhn & Zillmer, 2015; Shi, 2020b).

Another aspect of development is the so-called antilogos skill. Antilogos skill manifests itself as the realization, coordination, and reconciliation of alternative arguments that might lead to one's shift of perspectives to an opposing direction (Billig, 1987; Iordanou & Rapanta, 2021). Argumentation theorist Douglas Walton (1989) famously identified the two goals of argumentation which informed how antilogos strategies should be employed: one is to secure commitment from the opponent that can be used to support one's own argument and the other one is to weaken the argument of an opponent by identifying and challenging its weaknesses. Through participating in AWM, it has been found that students not only became more proficient in employing advanced counterargument strategies that directly reduced the force of the opposing argument by either critiquing its conclusions or undercutting its premises/reasoning (Kuhn et al., 2016; Mayweg-Paus et al., 2015; Papathomas & Kuhn, 2017;), but also gained facility in coordinating and reconciling contrasting arguments by constructing "*However*" arguments (Shi, 2019, 2020), achieving what others term the important ability of critical, integrative argumentation (Nussbaum, 2021) or dialectic argument (Rapanta & Felton, 2021).

Finally, meta-dialog (Krabbe, 2003) both as a socio-epistemic tool to regulate the process of argumentation and an important manifestation of argumentive competence develops through participating in AWM (Kuhn & Zillmer, 2015; Kuhn et al., 2013; Shi, 2020b). As students engage in argumentation with others by articulating, questioning, critiquing, and rebutting

arguments, they reveal, revise collaboratively, and become committed to certain norms and standards on which they base their construction and evaluations of each other's arguments. Students' meta-talks concentrate on the processes of argumentation that are distinguished from and govern their conversation about the content of the topic. With sustained involvement in argumentation, students engaged in more meta-talk not only with their same-side partners to meta-scaffold each other's understandings and reasoning (Zillmer & Kuhn, 2018), but also with the opposing party to check for mutual understandings and regulate the qualities of each other's arguments (Kuhn & Zillmer, 2015; Kuhn et al., 2013). Overtime, students' metatalks became more sustained and reciprocal. They were also charged with an investigative purpose as students insisted on getting a meta-level response from the interlocutor. This contrasted with descriptive meta-talk students most often involved with at the beginning of the program, which were always dismissed or only shallowly responded to by the opposing party (e.g., responded by a simple combative disagreement or simple clarification) (Kuhn & Zillmer, 2015).

The development of meta-talk signals more fundamental changes in students' metacognition and epistemological understandings. Meta-talk regarding argumentation is inherently epistemic because of the central role argumentation plays in the process of knowing through theory-evidence coordination (Macagno, 2016). It's been reported that accompanying students' advances in argumentive skills was their increased appreciation and understandings of the relevance and functions of evidence to arguments (Kuhn & Crowell, 2011; Shi, et al., 2019). In addition, students' epistemological understandings regarding the nature of knowledge and the process of knowing grew as a result of participating in AWM, alongside their advances in participating in meta-talks (Shi, 2020b). Advances in epistemological understanding underlies the development of intellectual values and dispositions to engage in argumentation. This

development at the motivational level is especially important because argumentative reasoning is cognitively demanding and effortful. Thus, for students to voluntarily participate in it outside of schools, they must be committed to its values (Kuhn et al., 2011).

1.5 Mechanisms of Development

The above cited research established the general efficacy of discourse especially (dialectic) argumentation as a pathway for developing individual argumentive skills. But the question regarding the underlying mechanisms still awaits a clear answer, especially given that dialogic skills are not explicitly taught in schools unlike expository writing.

In developmental psychology, Vygotsky and Piaget have long been the theorists most often turned to regarding developmental mechanisms. According to Vygotsky, all higher mental processes (as opposed to primitive, reflexive functions like involuntary attention) originate in the social plane in the process of an individual performing a task with other(s). Social practices can be interiorized/appropriated (Rodoff, 1995) by the individual to advance what one can do independently. During this sociocultural learning process, Vygotsky emphasized the critical concept of the "zone of proximal development" (ZPD) within which a more skilled other scaffolds the performance of a less capable one, bridging the gap between the current and the future ability levels of the less skilled. The process of internalization is not a simple mental duplication or accretion of what is originally social but entails a dialectical transformation process that will result in something qualitatively new at both the social and the individual planes.

In contrast to Vygotsky's emphasis on the process of internalization and the ZPD, Piaget (1976) offered another perspective regarding the mechanism underlying and the optimal

condition for social influence. Piaget focused on individual cognitive conflict as a primal factor in driving cognitive development (Piaget & Inhelder, 1969). In his perspective, children work as scientists trying to make sense of the surrounding physical and social environments. Development occurs as they strive to resolve discrepancies between their current understandings of the world and incoming new information. By altering their thinking to better fit the reality, children arrive at a qualitatively higher level of development or, in his words, an advanced level of equilibrium. Social factors can accelerate or retard children's progression along the hierarchies of cognitive development, but they are only secondary rather than determinate influencers.

In characterizing the most conducive (social) learning environment, Piaget paid special attention to peer-peer interactions. Differences in perspectives are revealed in discussions among similar ability partners that result in individual cognitive conflict. The individual is then prompted to work out a solution of this internal conflict, the resolution of which will lead to an advanced level of understanding. Peer-peer interaction is believed to be more likely to induce authentic reciprocal intellectual exchanges in the absence of power imbalance than adult-child interaction. In discussions, dialogic partners of comparable abilities can take on the role of support provider and receiver interchangeably and flexibly as needed, sometimes even at the meta-level (Zillmer & Kuhn, 2018), which scaffolds each other's learning (Halpern, 2022).

Forman (1987) insightfully commented that the difference between Vygotsky and Piaget's explanations of developmental mechanisms could be traced back to their characterizations of collaborative learning. Both of them noticed the parallels between the social and the individual practices. Vygotsky believed that they demonstrated that the individual processes derived from the joint social problem-solving. As to Piaget, however, he regarded such

similarities as evidence of both originating from the same central intra-psychological process while the individual struggles to work out internal cognitive conflicts.

1.6 The Apprenticeship Model

Numerous experimental studies including those cited above have established that dense engagement and practice with peers advances argumentation skill development, lending support to Piaget's view. Yet children and adolescents commonly engage in argumentive discourse in their daily lives not only with peers but also with older individuals possessing more advanced skills. This phenomenon has significant developmental significance, according to Vygotsky's theory. Children participate in discourses with older partners directly, but they also do so vicariously by overhearing it. How do the young who are exposed to more advanced argument strategies modeled by the more skilled interlocutor attend to and interpret them, eventually incorporating them into their own repertory to see how they fare? How and why do they do so? The answer is not straightforward nor sufficiently answered by Vygotsky. Rapanta and Felton (2021) ask, "How can learners learn how to argue effectively, when effective engagement in argumentation is a necessary part of such learning?" (p. 496). In a word, how can they appreciate, and go on to adopt, skilled argumentation strategies without already possessing the skill and understanding of purpose they entail? Fedyk et al., (2019; Fedyk & Xu, 2017) address this question by positing young children's possession of an epistemological theory of evidence that supports their learning, enabling them to recognize expertise and recognize its relevance to their goals.

Children don't often receive direct feedback with respect to argumentation strategies and must depend largely on observing the outcomes of others' strategies or of their own efforts. Do they infer different kinds of strategies, of either their own or of others, to be more or less successful and shape their own behavior accordingly? Presumably with time, strategies observed to be successful gradually make their way into the novice's repertory and begin to replace less successful ones.

A model that incorporates the multiple processes of observation, participation, and feedback (either direct or vicarious through observation) can be termed an apprenticeship model. The concept of cognitive apprenticeship has been applied to basic cognitive skills, most notably literacy (Scribner & Cole, 1981/2013), but not to our knowledge has it previously been explicitly adopted as a model of the acquisition of higher-order cognitive skills such as argumentation. The application is nonetheless consonant with the now widespread sociocultural perspective that emphasizes learning as apprenticeship into communities of practice (Rogoff, 1990).

1.7 The Present Study

In Vygotsky's (1987) classic portrayal of intellectual apprenticeship (though this is not a term he used directly), a more skilled individual participates in an activity with a less skilled one, and within this joint activity the latter's skills are practiced and develop, in a "zone of proximal development" (ZPD) until they can function independently. In embracing Vygotsky's view that a main goal of education is to develop higher mental functions, we structured the present learning experience to support students in adopting forms of argumentation they do not yet execute alone. During their joint activity, participants interact with a more skilled other who employs these

argumentation strategies, leading the less skilled individual to observe them in action and to recognize their effectiveness, with the task demands motivating the participant to try them.

Compared with existing studies on argumentive skill development (e.g., Hemberger et al., 2017; Kuhn & Moore, 2015; Kuhn, et al., 2016; Matos, 2021; Shi, 2019), the atypical characteristic of the present study is that repeated electronic discourse takes place both between two peer-pairs but also between a peer-pair and an older, more skilled discourse partner. The text-only medium allows substitution of the more skilled dialogic partner, absent the awareness of the target participant, thereby removing confounding social factors. The design includes a comparison condition whose participants engage in dialogs only with same-aged peers.

This hypothesis has received some initial support. One study made comparisons only at a local episodic level (Mayweg-Paus et al., 2015); two others (Halpern, 2022; Papathomas & Kuhn, 2017) showed effects after longer periods. Neither, however, took the important step we do here of showing that the effect extends to a face valid measure of skill in individual written argument, i.e., the traditional expository essay that is widely regarded as a crucial individual academic and life skill.

An additional new step is to employ an assessment tool for measurement of dialogic argument skills introduced by Zavala and Kuhn (2017) – a constructed dialog an individual is asked to produce in written form, depicting an argument between two skilled arguers. This solitary task allows assessment of an individual's understanding and skill with respect to the intrinsically interactive activity of discourse while remaining unaffected by an interlocutor's input.

We thereby seek to establish that any newly developed dialogic skill, while important, individuals can apply to and utilize successfully in performance of this critical individual skill.

In addition to its significance for educational practice, this demonstration will support an apprenticeship model by showing that in interaction with a more skilled other, the higher-order skills in question will have truly passed from the social to the individual plane.

Chapter 2: Method

2.1 Participants

Participants were 74 fifth grade students attending a public elementary school in a low-income rural county in the Northern region of China. Their ages ranged from 10 years 4 months to 11 years 6 months and 57% were female. They were starting their second semester of fifth grade when the study began. Their school was a large one containing ten fifth-grade classes of 70-80 students each, with two of the ten distinguished as accelerated classes based on first-grade entrance exam scores. One average fifth-grade class of 74 students was randomly selected to participate. Enrollment in the school followed the "enrollment in a nearby school" and "selection by housing" principles (Jin, Wang, & Huang, 2023; Ha & Yu, 2019), which means that the school exclusively enrolled school-aged children whose parents were homeowners within the nearby, predefined attendance zone. The school served a middle to upper-middle income neighborhood (relative to the overall income spectrum of the county), although the county itself was identified as a "national-level poverty-stricken county." Participants were all native Chinese-speakers. Almost all (98.2%) identified as ethnic Han. Written consent to participate was obtained from both students and their parents; no students or parents declined.

2.2 Design

Conditions. A total of 32 students were randomly selected to serve in an experimental condition. They were randomly assigned to dyads who held the same initial position on the topic to be debated and who remained intact during its discussion. The intervention included a series of electronic dialogs that took place sometimes with their peers and sometimes, unbeknownst to

them, with a college student from a group who had volunteered to play this role. The remaining 42 students, also randomly assigned to dyads who remained intact, served in a comparison condition. They participated in the same intervention but their intervention dialogs were restricted to peers. Students remained unaware which condition they had been assigned to. Participants in the two conditions underwent identical pretest and posttest individual assessments to be described.

Experimental manipulation. A **t**eam of 16 undergraduate students in their sophomore year from a tier-2 level university in the Northeast region of China partook virtually to implement the experimental manipulation. Unbeknownst to the younger participants, the undergraduates substituted as participants' peers in the experimental condition during two thirds of the electronic dialogues that participants took part in. It was expected that these undergraduate students, henceforth referred to as "experts" in argumentation (relative to the younger participants), would display more advanced argument strategies (in particular more advanced counterarguments and argument-evidence coordination strategies), as well as probe the younger participants to a greater extent about their own positions. To ensure that these experts would in fact demonstrate more advanced argument strategies, the researcher met with them in person for a three-hour session, introducing the dialogic argumentation activity as well as the online debate platform that was to be used. Dialogs were practiced among the group.

The intervention was described to the young participants as an argumentation and decision-making class. Students devoted four 40-minute class sessions per week to the class (except for two weeks when they had a mid-term exam and a national holiday break and thus had only two sessions for those weeks), for a total of 16 40-min sessions over a period of one and a half months. These class sessions normally would have been devoted to PE and Music subjects.

2.3 Intervention Procedure

Classroom setting. The first author served as the lead coach to implement the program. Two teachers from the school staff served as assistants. Their main duties were to circulate in the classroom while the lead teacher gave instructions to the whole class. The assistants ensured students were attentive, helped distribute materials, and helped in answering students' questions. To ensure the two assistants' interactions with students were consistent with the purposes of the study, the lead coach (first author) worked with them for one 60-minute session two days prior to the start of the intervention. During this session they reviewed activities students were to engage in and their purposes and viewed video clips of previous implementations of the intervention protocol. The two assistants remained unaware of the experimental manipulations and research questions.

Manipulation. During the dialogs of participants in the experimental condition, and unbeknownst to them, for three of each four dialog sessions (the first, second, and final), one of the adult experts substituted for the opposing-side peer dyad. During the other dialog session (the third), experimental dyads argued with an opposing-side peer dyad (also from the experimental condition). Participating students were able to remain blind to the experimental manipulation because the dialogs took place electronically and students were assigned pseudonyms. Dyads in the comparison condition only debated with opposing-side peer dyads.

Topics. Two topics from a set of 10 potential topics on which students' initial positions had been solicited were chosen for the intervention. The two were ones on which the class's chosen positions split roughly equally across opposing positions. The first topic was whether

teenagers should be allowed to drink soda or their soda consumption be restricted. The second topic was whether or not use of animals in research should be allowed. Students engaged with the soda topic first and then the animal research topic; the topics were so sequenced because the former was considered a more personal topic which would be easier for students to address and thus could introduce and familiarize them with the argumentation activities. The latter topic was more challenging and might serve to consolidate and extend relevant skills gained from engaging with the first topic.

Translation. One of the authors undertook translation of all materials into Chinese from their original English. All translated materials underwent the translation-backtranslation-compare-revision-translation process as many times as needed until no semantic difference existed between the back-translated and the original versions.

Intervention sequence. The dialogic argumentation intervention for both experimental and comparison conditions follows the rationale and framework presented by Kuhn and Crowell (2011), detailed most recently by Kuhn (2018); and implemented and reported on by others (Hemberger et al., 2017; Iordanou, 2010; Matos, 2021; Shi, 2019).

A sequence of eight sessions took place for each topic, each taking place during a class session of 40 minutes.

The sequence of activities for each topic were as follows.

Same-side team work (session 1). Students who shared the same position formed a team, who first convened in small groups of 4-5 students to generate, discuss, and evaluate reasons in support of their team's position. Groups were instructed to select one member as scribe to record reasons, compare them, and eliminate duplicates. Groups then assembled as a team to compare reasons and assemble a final set to display.

Opposing-side dialogs (sessions 2-5). Students formed dyads within their same-side teams. The composition of the dyads remained the same for each topic.

Text-only software was pre-installed and dialogs between pairs from the opposing team were conducted electronically via this software. Each dyad was assigned an animal name and instructed not to disclose any personal information during the dialogs and to engage in verbal conversation only with the same-side partner as they planned what responses to make to the opposing pair. After a few corrections during the first dialog, this instruction was adhered to.

Each dyad debated with a different opposing dyad at each of the four dialog sessions for the topic. Students were instructed to discuss and reach agreement with their same-side partner before responding electronically to the opposing pair. While awaiting responses from the opposing side, student pairs were asked to complete one of two written reflection activities, one asking them to reflect on one of the main arguments made by their opponents and their counterargument to it and the other asking them to reflect on one of their own main arguments, how it was countered, and their rebuttals to the counterargument. The two forms alternated across sessions; student dyads worked on one at each session and were reminded to discuss with their partner and agree before writing. Students' dialogs were monitored by the lead coach via a master computer. If students' discussions were detected to go off-topic, a reminder for them to return to the topic was given. Otherwise, no feedback was provided.

During each of the dialog sessions, dyads were provided factual information related to the topic in the form of short Q&A items. They were advised that this information might be helpful to them in making their arguments but they were not required to use it. They were able to choose questions and access their answers during the periods while they were awaiting responses from the opposing dyad. Across the four dialog sessions, eight such pieces of information were made

available, together representing the four potential functions such information might serve as evidence.

1. Potential evidence that might be used to support one's favored position (M+)
2. Potential evidence that might be used to weaken the opponent' position (O-)
3. Potential evidence that might be used to support the opponent's position (O+)
4. Potential evidence that might be used to weaken one's own position (M-)

This sequence of presentation parallels the hierarchy of cognitive demands each type of evidence poses (Hemberger et al., 2017; Kuhn et al., 2016; Shi, 2019). It started with the easiest, support-own, and then progressed to the more challenging, weaken-other type – more challenging because to envision its use, students needed to shift their attention from their own position to the opponent's position. The most challenging types of evidence, presented last, were the two that are incongruent with one's favored position (support-other and weaken-own), the inclusion of which warrants effort to integrate and reconcile with one's own position.

Students were also encouraged to ask questions of their own whose answers might help them in their argumentation. Most students undertook the task of searching for answers to their self-generated questions on the internet themselves, but they were provided help if needed. Once answered, the question and answer were made available to all students.

As the end of each dialog session, students were asked to rate their opponents' performance ("How good an arguer was your opponent today?") on a scale of zero to 100. The purpose of this activity was to assess whether the manipulation of more versus less capable interlocutors was successfully perceived by participants.

Showdown debate (session 6). Prior to this "showdown" session, each team was asked to select 15 members to be the representatives of their team in a showdown debate. During the

debate, representatives took turns to occupy one of the two "hot seats" placed in the middle of the classroom and verbally debated one another. Each round of debate between the hot-seat occupants lasted three minutes, and then a new pair took their place. During each round, a "huddle" of one minute could be called either by one of the "hot seat" speakers themselves or by other members of their team. In total, eight rounds of debate took place at the topic 1 showdown debate and 10 rounds at the topic 2 showdown debate.

Showdown debrief (session 7). All showdown sessions were videorecorded and subsequently transcribed by one of the authors, who constructed an argument map for each one. It was divided into idea units consisting of a claim with its supporting reasons and evidence. Then, each idea unit was assigned a functional code in relation to the utterance immediately preceding it, following a coding scheme originated by Felton and Kuhn (2001) and further developed in subsequent studies by a number of researchers. Effective codes included counterarguments and rebuttals that aimed to weaken the opponent's arguments, in addition to the functional use of evidence to address a claim and attempts to elaborate and clarify in response to an opponent's queries. Ineffective codes included statements unconnected to what preceded them and/or concentrated exclusively on elaborating one's own arguments without attention to the opponent's position. Each effective code was awarded one point while ineffective code received a negative point. Evidence-based statements received one additional point.

During the debrief session the lead coach led students through an argument map, which was presented to the whole class in a series of PowerPoint slides. During the presentation, students were encouraged to voice any doubts/disagreement with respect to the assigned codes. At the debrief session for topic 1, only one student questioned the assignment of an idea unit to the category of *counter-undermine* (the highest level of counterargument, which challenged the

opponent's reasoning). At the debrief session for topic 2, more than five students either questioned or voiced alternative perspectives regarding the categorizations of the idea units, one of which resulted in a brief whole-class discussion and change of scores. At the end of the debrief session, scores were summed for each team and a winning team announced.

Individual essay (session 8). The final activity for each topic was the only one during which students worked individually. The task was to write an individual essay on the topic. The essay was framed as a "a letter to a newspaper editor" for the purpose of evaluating and comparing the opposing positions on the topic. A list of 12 pieces of potential evidence related to the topic (including the eight pieces of evidence students encountered during the earlier sessions) was available to students. Students were allowed the entire class session of 40 minutes to complete their essays. No word limits were suggested to them.

2.4 Pretest and Posttest Assessments

Constructed dialog. One week prior to the start and one week after the completion of the intervention, students from both conditions completed the same solitary dialog construction task administered by their classroom teacher as a class assignment and not connected to the intervention, whose coaches were not present during its administration. The task, adapted from Zavala and Kuhn (2017) and later Shi (2020a), assessed students' understanding of and skill in argumentation. The task asked students to individually construct an argumentive dialog between two expert arguers on the topic of whether adolescents who have committed serious crime should be tried in adult or juvenile court. This topic was not addressed during the intervention nor had it been previously discussed by students in the school, as confirmed by their teachers. A set of eight pieces of potential evidence relevant to the topic, balanced in terms of their relative

functions in relation to both sides (i.e., the M+, O-, O+, and M- functions), were made available but students were told they were not required to use them when constructing their written dialogs. The dialog was described as taking place between two skilled arguers.

Most students completed the task within 25 minutes and all the students completed it within the assigned 40-minute period.

Essay. Occurring on a different day but in the same week and following completion of the posttest dialog construction task, an individual posttest essay writing task was conducted in students' classrooms. The purpose of this task was to assess transfer of skills from the dialogic to a conventional individual essay context. All participants from both conditions completed the task in a class session of 40 minutes. Only their head teacher, and not any of the intervention coaches, was present in the classroom. Students were asked to write an individual essay regarding the topic not addressed in the intervention of whether adolescents who have committed a serious crime should be tried in adult or juvenile court. The evidence items for the dialog task were available, along with the same instructions.

Once students had completed both assessments, they were more fully debriefed by their teacher regarding the purposes of the activities and of the research being conducted.

Chapter 3: Results

Coding schemes were employed to apply to each of the assessment instruments, individually constructed dialogs and traditional individual essays, both based on an unstudied, transfer topic not addressed during the intervention.

All participants' identification information and condition assignments were removed prior to coding. In total, 145 constructed dialogs (one pretest and posttest dialog for all but two students who missed one or both due to medical absences*)* and 70 essays, except for four students who were absent due to conflicting activities, were coded.

Both dialogs and essays were coded by an author and a research assistant who was familiar with the intervention paradigm but not familiar with the research questions, design, or implementation of the study; 30 % of dialogs (and 25% of essays) were randomly selected and independently coded by the two coders.

3.1 Coding of Constructed Dialogs

Idea units. Dialogs were divided into idea units (most often a turn in the dialog), and idea units were then coded in two ways. The first-order and basic level classified the dialogic function the unit served in relation to the immediately preceding unit. This coding scheme was adapted from ones described by Halpern (2022), by Shi (2020a) and in previous versions by Crowell and Kuhn (2014) and Zavala and Kuhn (2017). Its validity was supported by expected correlations between students' argumentative thinking assessed by this coding scheme and other related constructs involving argumentative thinking such as individual argument production, epistemological understanding (Shi, 2020b) and mastery of content knowledge (Iordanou et al., 2019).

Idea units were defined as a claim with its supporting reasons and/or information serving as evidence, intended to convey a single point (Iordanou & Rapanta, 2021). A differentiation was made between on-topic and off-topic idea units, which were not directly related to the issue to be debated. Off-topic idea units digressed from the focal topic and did not include a claim with respect to it (e.g., how to build juvenile courts more efficiently). Off-topic idea units reflect weakness in understanding the purposes of argumentation and/or in proactive executive control abilities (Nussbaum & Asterhan, 2016). Only on-topic idea units were included in the present analyses. Percentage agreement between the two coders in classifying on- and off-topic idea units was 96% (Cohen's Kappa = 0.90, $p < .001$).

Next each on-topic idea unit was classified into one of the categories presented in Table 1. Unconnected idea units bear no discernable connection to the unit immediately preceding them and like off-topic units reflect failure to fully understand the purposes and functions of argumentation. Percentage agreement between the two coders in classifying idea units into one of the categories in Table 1 (questioning, meta-talk, the four types of countering, and concession) was 90% (Cohen's Kappa = 0.88, $p < .001$).

Table 1: Categories and Examples of Types of Idea Units in Analytical Scheme for Coding Idea Units in Constructed Dialogs.

Category	Description	Examples
Questioning	A statement requesting a response from the interlocutor, typically ending with a question mark	What if there is no juvenile court? Should an adolescent who committed a very serious crime like murder also be given a lenient sentence?
Meta-talk	A statement about the dialog itself	You didn't understand what the evidence said. I don't want to talk about this; let's go back to discuss juvenile courts.

Counter-disagree	The weakest type of counterargument, rejecting the interlocutor's preceding statement without justification	I don't agree. You are wrong.
Counter-alternative	A weak form of counterargument, criticizing the interlocutor's preceding claim indirectly by introducing an alternative but not necessarily incompatible argument	*(Interlocutor: If the government doesn't have to open a separate juvenile court, they can save a lot of money.)* But juvenile court is more suitable for adolescents because their brains haven't developed fully.
Counter-critique	A strong form of counterargument that directly criticizes the interlocutor's conclusion, typically by pointing out its negative consequences	*(Interlocutor: I think adolescents should receive a lenient sentence.)* But this will only make them more aggressive in the future.
Counter-undermine	The strongest type of counterargument, criticizing the interlocutor's argument by either rejecting one of its premises or showing the conclusion doesn't follow from the premises	*(Interlocutor: The staff in juvenile courts might not understand adolescents' thinking because they are already grown-ups.)* But everyone grows from an adolescent to an adult, how come they don't understand? But everyone grows in different environments and who knows what environments these adolescents grew up in.
Concession	Acknowledgement of merits of the interlocutor's position or weaknesses of one's own position.	*(Interlocutor: In juvenile detention, their criminal record will be sealed upon their release and not impact their future.)* This might be good. But …

Each of the types in Table 1 play a functional role in argumentation. Questioning plays an essential role in initiating, probing, and deepening a dialog. It invites elaborations and prepares the way for counterargument. Meta-talk plays a particularly important role in deepening dialog through reflection on and questioning of the dialogic exchange itself (e.g., whether an argumentative move was acceptable or whether the discussion adhered to the topic) rather than its content. Such reflection paves the way for a deeper understanding of the purpose and value of argument.

Countering and rebutting constitute the core of argumentative dialog because they are the key strategies employed to weaken the force of the interlocutor's arguments and to defend and restore the strength of one's own claims (Walton, 1989). Four forms of counterarguments can be identified, of varying power. *Counter-disagree* is the weakest form; it simply expresses a disagreement with the interlocutor's claim. *Counter-Alternative* goes beyond simple disagreement by offering an alternative claim, leaving the interlocutor's argument unaddressed. But in fact, the two may not contradict one another.

Counter-critique is a stronger form of counterargument because it weakens the strength of the interlocutor's argument by directly and explicitly criticizing its conclusion usually by pointing out undesirable consequences.

Counter-undermine is regarded as the strongest type of counterargument because it attacks the reasoning underlying the interlocutor's argument. This is achieved either by rejecting the explicit or implicit premise on which the interlocutor's argument is grounded or by challenging the link between the premise and the conclusion the interlocutor draws.

Concession is the first step toward reconciliation of opposing claims in the sense that it acknowledges either merits of the opposing position or weaknesses of one's own position. It does more than simply capitulate without justification (e.g., "Okay, I agree with you").

Evidence coding. A second level of coding of idea units was applied to examine students' ability to use evidence to strengthen their arguments. An idea unit was coded as evidence-based if it included a reference to evidence. Evidence could be drawn from the information made available to students (i.e., shared evidence) or it could consist of commonly accepted knowledge students drew from their personal knowledge (i.e., personal evidence). A distinction was made between functional and non-functional uses of evidence. Occasionally, students merely inserted a piece of information into their constructed dialogs without linking it to any arguments they intended to make, and sometimes they mischaracterized the information, invalidating its use as evidence. In each of these cases, the evidence-based units were classified as non-functional. To qualify as a functional use of information as evidence in relation to a claim, the connection to and the function of the information in relation to the claim must be made clear. Sometimes a connection was attempted but unsuccessful, in which case the unit was classified as non-functional use of evidence (e.g., "*According to report, teens were involved in one quarter of violent crimes in the USA over a 25-year period beginning in 1990*, *this shows the importance of having adult courts."*). Percentage agreement between the two coders in coding and categorizing idea units as functional evidence-based units was 95% (Cohen's Kappa = 0.90**,** $p < .001$), only functional use of evidence was included in the present analysis.

***However* unit-pairs**. These constitute a third level of coding. Rather than a descriptor of a single idea unit, they are based on the nature of the connection between two adjacent idea units. Three criteria must be satisfied for two adjacent idea units to qualify as a *However* pair. First,

one must appear immediately after the other and they must be connected to one another with respect to their content. Second, they must be *explicitly* connected to one another, by a conjunction such as "however," "but," or "although." Third, the two statements must oppose one another, i.e., they must argue in opposing directions, thus serving opposing argumentive functions. The functions to be introduced later (see Table 5) as the primary ones that were applied to coding of idea units in essays (support-own, weaken-other, support-other, weaken-own) were applied to any two adjacent units, and the unit-pair was classified as a *However* pair if the pair was of any of these types*:* support-own and weaken-own; support-other and weaken-other; support-own and support-other; weaken-own and weaken-other. Examples of successful and unsuccessful *However* pairings appear in Table 2. Only successful ones were included in analyses. For the coding of *However* pairs, percentage agreement between the two coders was 96% (Cohen's Kappa = 0.92, $p < .001$).

Table 2: Examples of Successful and Unsuccessful However Unit-pairs.

Successful *However* pairing	Adjacent statements are explicitly connected in content and syntax and reflect opposing argumentive functions	*Although adolescent's future will not be destroyed if their records are sealed, this is unfair to the victims who got hurt.* *Though juvenile courts are more expensive to run, teens get better treatment there so it's worth the money.*
Unsuccessful *However* pairing	Adjacent statements not connected in content.	Judges in juvenile courts give shorter sentences, but adult courts cost less money.

Two examples of coded constructed dialogs, one from the experimental condition and the other the comparison condition, are shown in Table 3 and Table 4. Statements coded in one of the successful categories appear in green.

Table 3. Example of Constructed Dialog from the Experimental Condition at Posttest.

Turn	Interlocutor	Contribution
1	LP	I think they should be tried in juvenile courts because juvenile courts give less punishment and shorter sentences, and this will reduce the pressure on young people.
2	WH	But to establish juvenile courts, the government will spend a lot of money, manpower, and resources, which is not worth it.
3	LP	Although juvenile courts cost a lot of money, they can save an adolescent physically and mentally. According to research, people's brain does not mature until they are 20 years old. Many teenagers break the law on impulse
4	WH	Although teenagers are immature in their thinking, they have to be responsible for what they have done. If they are not punished for their crimes, some of them will make mistakes again and again without repentance and never improve.
5	LP	But teenagers are at risk of being beaten up in adult prisons, and if they are beaten, it will become a psychological shadow for them for life.
6	WH	But can you guarantee they won't be beaten in juvenile prisons? Adult prisons comprise of mostly adults, so there will be no irrational behaviors.
7	LP	If they had been rational adults, they would not be in prison. Juvenile prisons are filled with young people under the age of 18 who are peers and can get along well with each other.

Table 4. Example of Constructed Dialog from the Comparison Condition at Posttest.

Turn	Interlocutor	Contribution
1	LP	Teenagers who commit crime should be tried in juvenile court.
2	WH	Why??

3	LP	Because in recent years, teenagers have also committed violent crimes such as murder.
4	WH	I don't agree because they have already committed murder which breaks the bottom line of juvenile courts.
5	LP	But teenagers face a very high risk of being beaten in adult prisons.
6	WH	This is how to keep young people in check.
7	LP	The criminal records of young people serving their sentences in juvenile courts will be sealed upon their release, without impacting their future.
8	WH	But this is unfair.

3.2 Coding of Essays

As was the case for their constructed dialogs, students' essays were first segmented into idea units which were then classified into on- or off-topic idea units. Acceptable percentage agreement between the two coders in classifying on- and off-topic idea units was 96% (Cohen's Kappa = 0.86, $p < .001$). Only on-topic idea units within an essay were further coded.

Next, a functional coding scheme was applied to categorize each on-topic idea unit into four categories based on the argumentive functions they serve in relation to the position being considered (here, whether adolescents should be tried in adult or juvenile court). This coding scheme was first used by Felton and Kuhn (2001) and subsequently elaborated and used in further studies. Table 5 presents definitions and verbatim translated examples from students' essays of each of the functional types. Four types were identified. They were *support-own* (M+) type that functioned to support the writer's position, *weaken-other* (O-) that functioned to challenge the opposing position, *support-other* (O+) that functioned to acknowledge merits of the opposing side, and *weaken-own* (M-) type that functioned to acknowledge weaknesses of

one's own side. Percentage agreement in classifying each idea unit into one of the four functional types was 93% (Cohen's Kappa 0.89, $p < .001$).

Table 5. Examples of Functional Idea Units in Analytical Scheme for Coding Idea Units in Essays.

Argumentive Functions	Definitions	Examples
Support my own (M+)	Statements functioning to support one's own position	Judges in juvenile courts can understand adolescents better (*pro juvenile court*) Adult courts are fairer for everyone (*pro adult court*)
Weaken other (O-)	Statements functioning to critique and thereby weaken the opponent's position	Adolescents will be bullied by adults in adult jail which will increase their hatred for the society (*pro juvenile court*) Teens cannot get enough punishment and learn their lessons in juvenile court (*pro adult court*)
Support other (O+)	Statements functioning to acknowledge strengths of the opponent's position	Adolescents can learn their lessons by having a longer sentence (in adult court) (*pro juvenile court*) Adolescents will not waste too much time in jail if tried in a juvenile court (*pro adult court*)
Weaken my own (M-)	Statements functioning to acknowledge weaknesses of one's own position	Juvenile court costs more money (*pro juvenile court*) Adult courts may give longer sentences (*pro adult court*)

In addition to the basic coding of idea units, the additional coding of evidence use and *However* pairs described earlier for coding dialogs was also conducted for essays. The same coding criteria described earlier were followed. Percentage agreement in coding and categorizing evidence-based idea units was 96% (Cohen's Kappa = 0.93, $p < .001$). Only functional uses of evidence were included in the present analysis. Percentage agreement in identifying and

classifying *However* pairs in essays was 95% (Cohen's Kappa = 0.87**,** $p < .001$). Only successful *However* pairs were included in the present analysis.

3.3 Experimental Manipulation Check

To assess whether the experimental manipulation was implemented successfully, both the ratings participants assigned to their interlocutors at the end of each electronic dialog session and the actual dialogic contributions of experts were evaluated. The former aimed to examine whether the manipulation was perceived by the participants as was intended to by the researchers. The latter intended to characterize experts' actual dialogic behaviors.

Dialogic partner performance ratings. To quantitatively evaluate and compare the performance of expert partners against that of peer partners, within the experimental condition, we conducted a dependent sample t-test to assess whether participants in general gave higher ratings to their expert partners than peer partners on a scale of 0 to 100. For their peer partners' performance, participants' ratings ranged from 31 to 89 with a mean rating of 60.52 (SD = 12.93), while for their expert partners' performance, participants' ratings ranged from 47 to 99 with an average rating of 75.14 (SD = 6.21). A dependent samples t-test revealed that the mean of the difference scores between the (average) ratings participants gave to their expert partners and peer partners was significantly different from zero, indicating participants in general regarded their expert partners as performing better than their peer partners , $t(15) = -4.44, p < .001$.

As an additional verification step, we also implemented an independent samples t-test to compare, across conditions, whether experimental and comparison students gave similar ratings

to their peer dialogic partners. Since experimental students only debated with peer partners during session 3 of each topic circle, for both conditions, we only selected participants' ratings from these two sessions for analysis to equate the sample sizes for both conditions. For experimental students, their ratings for peer dialogic partners' performance ranged from 31 to 89 with a mean of 60.52 (SD = 12.93) and for comparison students, their ratings ranged from 35 to 88 with an average rating of 58 (SD = 11.44). An independent sample t-test revealed that this difference between the two means was not statistically significant from zero, $t(35) = -0.63, p = 0.53$, suggesting participants from the two conditions gave comparable ratings to their peer dialogic partners.

Expert dialogic contributions. To get insight into how experts interacted with the younger participants in real electronic dialogs, experts' statements in dialogs were coded and analyzed using the same coding scheme developed for coding participants' constructed dialogs. The results indicated that experts' dialogic contributions were characterized by high proportions of probing units attempting to regulate the younger participants' moves by (meta-) questioning and units categorized as advanced counterarguments. Specifically, across the 16 experts, a mean of 33% (SD = 0.19) of their statements were classified as questioning and 20% (SD = 0.14) as meta-talk units. Another 30% (SD = 0.18) of their contributions constituted as counterargument units with 93% (SD = 0.12) of them categorized as direct counters (i.e., counter-critique or counter-undermine). Regarding evidence use, on average, 17% (SD = 0.12) of experts' idea units were successfully supported by evidence and all experts advanced at least one functional evidence-based unit in each electronic dialog sessions.

We turn now to the usage of these categories in the constructed dialogs of participants themselves.

3.4 Statistical Analysis of Constructed Dialogs

The number of individually constructed dialogs analyzed at pretest was 32 for the experimental condition and 41 for the comparison condition. At posttest numbers were again 32 for experimental condition but 40 for the comparison condition (due to absence). The unit of analysis was the idea unit in a participant's individually constructed dialogs. The main outcome variables were the frequencies of occurrence of the seven types of units described earlier in Table 1, frequencies of evidence-based units, and frequencies of the *However* pairs described earlier and in Table 2.

The main statistical analysis consisted of negative binomial analysis to answer the research question of whether experimental students who argued with a mixture of less and more capable partners will show greater skill gains than comparison students who argued only with their similar ability peer pairs. Students' pretest performance was first analyzed to establish equivalence of skills before the intervention commenced. The independent variable for both pretest and posttest analyses was a between-subject two-level factor (experimental, comparison). For posttest analyses, relevant pretest performance was controlled for by involving it as a covariate.

Negative binomial analysis was chosen as the more appropriate tool for analyzing the present data because the dependent variables were count data of non-negative values. Standard regression techniques (e.g., least squares regression) are inappropriate to model the current data because they require the distribution of the dependent variable to be normal. Poisson regression analysis may be favored in the case of count data which relaxed the normal distribution assumption of regular regression models by applying a log transformation on the dependent variable. However, the standard Poisson regression model is restricted by another assumption

that requires the conditional mean and variance of the dependent variable being equal (i.e., mean and variance taking into account the covariate(s) in the model). When this assumption is violated, a negative binomial regression is favored which relaxes the assumption of equal conditional mean and variance by adding a random error term to the standard Poisson model and has more power compared to other alternatives (e.g., the quasi-Poisson analysis) (Hilbe, 2011). Since many outcome variables within the current dataset were afflicted by the problem of overdispersion (i.e., the conditional variances of the dependent variables were larger than their respective conditional means), standard poison analysis was inappropriate for them. When the conditional means and variances of the dependent variable are equal (negative binomial regression and Poisson regression producing the same outcomes), negative binomial regression is the most appropriate technique to apply to the current dataset. This assumption was supported by Goodness-of-fit statistics comparing Poisson regression models and their associated negative binomial models that indicated negative binomial analysis to be a better fit for the present dataset. Thus, we report negative binomial regression analyses results in the subsequent sections.

To answer the research question of whether students within each condition made significant progressions from pretest to posttest in argumentive skills, a second set of analyses was carried out comprising of dependent samples t-tests. Wherever the assumption of normally distributed data was violated, the Wilcoxon signed-rank test, a nonparametric alternative to dependent samples t-test, was carried out in place of the t-test.

Two additional individual-level analyses were carried out to supplement the negative binomial analyses and dependent samples t-tests, which captured only group-level trends. These identified the proportion of students who ever included a particular type of idea unit in their dialog. This analysis is useful in assessing whether any advantage exhibited by the experimental

group over the comparison group at posttest was due to the exceptional performance of a limited group of students. (Pearson's Chi-square Tests of Independence were carried out to answer this question, with Fisher's Exact Test of Independence substituted wherever expected value fell below the criterion of five.) Then, within each condition, to examine whether the intervention led to significant increases in the percentages of students who ever used a particular argumentative strategy from pretest to posttest, McNemar's Test was employed, a non-parametric test to analyze paired count data with a dichotomous independent variable. (Wherever the expected frequencies in the two discordant cells fell below the criterion of five, the exact binomial test was performed in place of McNemar's test.)

Idea units. At pretest, the experimental group generated a mean of 3.47 (SD = 2.24) on-topic idea units in their constructed dialogs and the comparison group a mean of 2.68 on-topic idea unit (SD = 2.15), a nonsignificant difference ($\hat{\beta} = 0.26, p = 0.14, IRR = 1.29$). As shown in Figure 1, at posttest, mean idea units in experimental students' dialogs increased to 5.72 (SD = 2.25) idea units, a significant change from pretest, $t(31) = -4.54, p < .001$; by comparison, comparison students' mean idea units increased to 3.98 (SD = 2.15), also a significant change from pretest, $t(39) = -2.82, p = 0.01$. After adjusting for pretest performance, a negative binomial regression analysis revealed that at posttest, experimental group dialogs contained 1.33 times more idea units than those of comparison group students, a significant difference ($\hat{\beta} = 0.29, p = 0.02$).

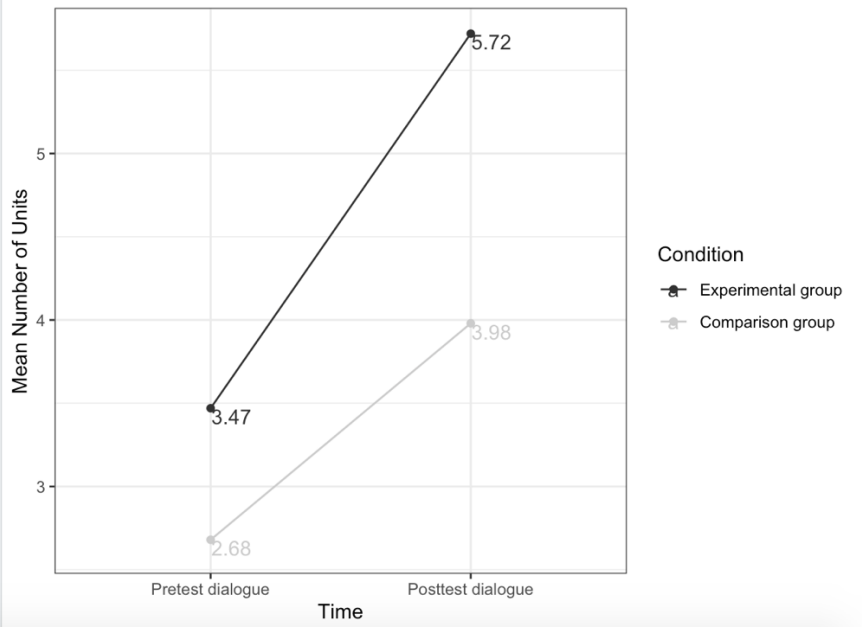


Figure 1. Total On-topic Idea Units in Constructed Dialogs by Condition and Time.

Question frequencies. Mean numbers of occurrences of idea units in the question category at pretest and posttest appear in Figure 2. At pretest, the two conditions generated comparable frequencies of question units (Means = 0.09 and 0.05, SDs = 0.30 and 0.22 for experimental and comparison conditions, respectively), a non-significant difference ($\hat{\beta} = 0.65, p = 0.48, IRR = 1.92$). At posttest, the two conditions again showed comparable numbers of question units in their dialogs (Means = 0.34 and 0.35, SDs = 0.60, 0.86 for experimental and comparison conditions, respectively.) As shown in Table 6, row 1, after controlling for pretest performance, the two conditions did not differ significantly in frequencies of question units in their posttest dialogs ($\hat{\beta} = -0.02, p = 0.97, IRR = 0.98$).

At the individual level, at pretest, 9% of participants from the experimental condition and 3% of participants from the comparison condition contributed to the occurrences of question units, a nonsignificant difference, $p = 0.65(Fisher's\ Exact\ Test)$. At posttest, 28% of participants from the experimental condition and 20% of participants from the comparison

condition did so, again a nonsignificant difference, $\chi^2(1, 72) = 0.65, p = 0.42$. Still, less than a third of students ever included questioning of the interlocutor as part of their constructed dialog, compared to the experts, among whom a third of all their contributions were questions.

Both conditions, however, showed significant improvement from pretest to posttest in the frequencies of question unit contained in constructed dialogs, in the experimental condition, $z = -2.09, p = 0.04$, and comparison condition, $z = -2.26, p = 0.02$. Both conditions also saw a marginally significant increase in the number of students contributing to such occurrences, $\chi^2(1, 32) = 3.6, p = 0.06$ for the experimental condition and $\chi^2(1, 40) = 3.6, p = 0.06$ for the comparison condition.

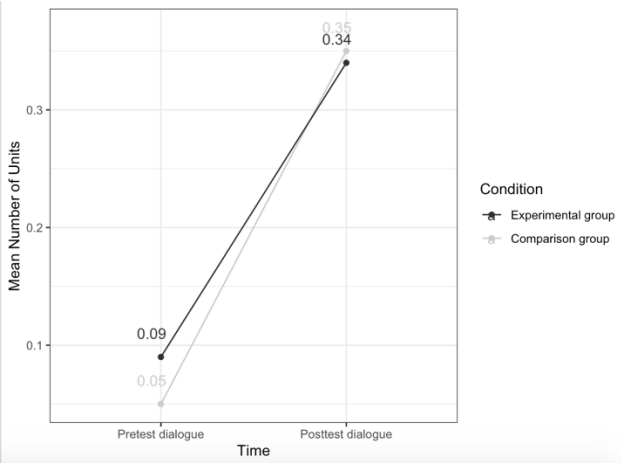


Figure 2. Question Frequencies in Constructed Dialogs by Condition and Time.

Table 6. Estimation Results of Negative Binomial Regressions on Posttest Dialog Occurrences.

Response Variable	$\hat{\beta}$	Estimates	Std Err of $\hat{\beta}$	Conf. Int. of $\hat{\beta}$	IRR[c]	Sig.
Question	Intercept	-1.25	0.34	(-1.97 to -0.61)		<.001
	Condition[a]	-0.02	0.50	(-1 to 0.95)	0.98	0.97
	Pretest[b]	1.44	0.69	(0.1 to 2.87)	4.23	0.04
Meta-talk	Intercept	-3.33	-5.20	(-5.20 to -2.11)		<.001
	Condition	1.44	-0.04	(-0.04 to 3.36)	4.20	0.08
	Pretest	2.20	0.28	(0.28 to 3.68)	9.04	.007
Indirect counters	Intercept	-0.70	0.25	(-1.22 to -0.25)		.005

	Condition	0.06	0.31	(-0.57 to 0.68)	1.06	0.85
	Pretest	0.36	0.16	(0.03 to .67)	1.44	0.02
	Intercept	-0.77	0.24	(-1.27 to -0.34)		.001
Counter-critique	Condition	0.76	0.28	(0.22 to 1.34)	2.15	**.007**
	Pretest	0.39	0.31	(-0.25 to 0.97)	1.5	0.21
	Intercept	-1.74	0.40	(-2.61 to -1.04)		<.001
Counter-undermine	Condition	1.86	0.46	(1.01 to 2.82)	6.40	**<.001**
	Pretest	-1.0	0.75	(-2.84 to .25)	.37	0.18
	Intercept	-2.35	0.50	(-3.52 to -1.51)		<.001
Concession	Condition	1.86	0.55	(0.88 to 3.09)	6.34	**<.001**
	Pretest	0.42	0.30	(-0.23 to 0.96)	1.52	0.16
	Intercept	-0.26	0.25	(-0.76 to 0.22)		0.29
Unconnected	Condition	-0.59	0.40	(-1.40 to 0.18)	0.55	0.14
	Pretest	0.15	0.18	(-0.21 to 0.50)	1.16	0.41

[a] condition = 0 which was the peer-only group that was taken as the reference group.
[b] pretest performance on various outcome variables were taken as covariate in respective models.
[c] the incidence rate ratios.

Meta-talk frequencies. Frequencies of idea units in the Meta-talk category at pretest and posttest appear in Figure 3. At pretest, experimental students on average generated 0.03 (SD = 0.18) meta-talk units compared to 0.05 (SD = 0.22) such units on the part of comparison students, a no-significant difference ($\hat{\beta} = -0.45, p = 0.72, IRR = 0.64$). At posttest, experimental students generated 0.19 (SD = 0.40) meta-talk unit while comparison students again produced 0.05 (SD = 0.22) such units in their constructed dialogs. After controlling for pretest performance, a negative binomial regression analysis revealed that the condition difference at posttest did not reach statistical significance ($\hat{\beta} = 1.44, p = 0.08, IRR = 4.20$, Table 6).

Given these low means, it is more relevant to identify the number of students producing them. At pretest, 1 (3%) student from experimental group and 2 (5%) students from comparison group generated meta-talk unit at least once, a nonsignificant difference, $p =$ 1.0 ($Fisher's exact\ Test$). At posttest, 6 (19%) students from the experimental group and 2 (5%) students from the comparison group did so, again a nonsignificant difference, $p =$

0.13 ($Fisher's exact\ Test$). In contrast, among the experts, 20% of all their statements were meta-statements.

Despite a lack of condition difference, experimental students did improve significantly from pretest to posttest both in terms of the frequencies of meta-talk units, $z = -2.09, p = 0.04$, and the number of students contributing to these occurrences, $\chi^2(1, 32) = 5, p = 0.03$. In contrast, comparison students demonstrated no change in either respect.

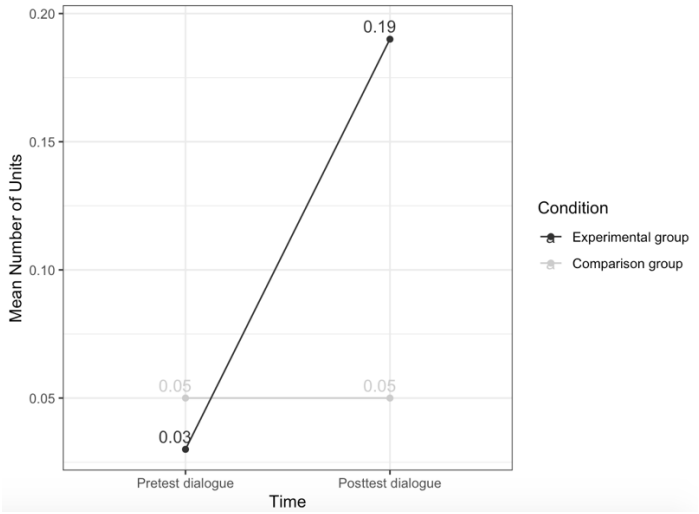


Figure 3. Meta-talk Frequencies in Constructed Dialogs by Condition and Time.

Indirect-counter frequencies. The two weak, least effective counterargument types, counter-D and counter-A, were combined into one category referred to as indirect counterargument types. Both counterargument efforts are unproductive. In both, the goal of weakening the opposing side's arguments is unsuccessful due to failure to directly engage with them, either offering no argument or substituting an unrelated one. Still these attempts can be seen as reflecting some progress in their effort to address and challenge the opposing position. As seen in Figure 4, at pretest experimental students on average generated 0.28 (SD = 0.52) idea units in the indirect-counter category in comparison to 0.56 such unit of comparison students (SD = 0.90), a non-significant difference ($\hat{\beta} = -0.69, p = 0.11, IRR = 0.50$). At posttest, the

mean number of indirect-counter units generated by experimental students was 0.59 (SD = 0.89); the mean number of indirect-counter unit generated by comparison students was 0.65 (SD = 0.76). According to Table 6, after controlling for pretest performance, a negative binomial analysis indicated that posttest condition difference was not statistically significant ($\hat{\beta} = 0.06, p = 0.85, IRR = 1.06$).

At pretest, the percentage of students contributing to the occurrences of indirect counter unit was 25% for the experimental condition and 37% for the comparison condition, a nonsignificant difference, $\chi^2(1, 73) = 1.12, p = 0.29$. At posttest, 47% and 45% of students from the experimental and comparison conditions did so, respectively, again a nonsignificant difference, $\chi^2(1, 72) = 0.03, p = 0.87$.

Within the experimental condition, students demonstrated significant change from pretest to posttest in terms of both frequencies of indirect counter units, $z = -1.94, p = 0.05$, and numbers of students contributing to these occurrences, $\chi^2(1, 32) = 3.77, p = 0.05$. No significant change over time was observed on the part of the comparison students, in terms of the frequencies of occurrences, $z = -0.36, p = 0.72$, and the percentages of contributing students, $\chi^2(1, 40) = 0.53, p = 0.47$.

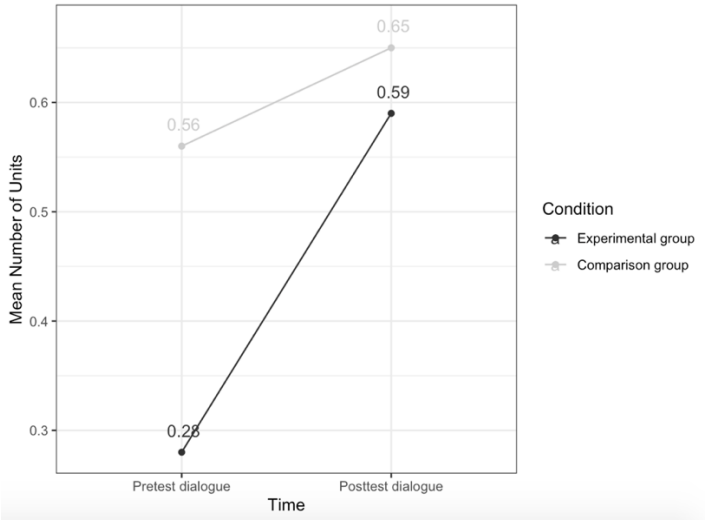


Figure 4. Indirect Counter Frequencies in Constructed Dialogs by Condition and Time.

Counter-C frequencies. The mean numbers of idea units categorized as Counter-C at pretest and posttest appear in Figure 5. As shown, at pretest, experimental and comparison students generated comparable numbers of Counter-C units (Means = 0.22 and 0.17, SDs = 0.42 and 0.38 for experimental and comparison students, respectively), a nonsignificant condition difference ($\hat{\beta} = 0.25, p = 0.64, IRR = 1.28$). At posttest, the mean number of idea units in the Counter-C category was 1.09 (SD = 1.12) for experimental students while 0.5 for comparison students (SD = 0.72). After controlling for pretest performance, as shown in Table 6, at posttest, experimental students generated an expected 2.15 times more counter-C units than comparison students, a significant difference ($\hat{\beta} = 0.76, p = .007$).

With regards to the numbers of students contributing to the occurrences of Counter-C units, at pretest, 22% of experimental students and 17% of comparison students did so, a nonsignificant difference, $\chi^2(1, 73) = 0.27, p = 0.61$. However, at posttest, 69% of experimental students compared to 40% of comparison students did so, a significant advantage on the part of the experimental students, $\chi^2(1, 72) = 5.90, p = 0.02$. (These percentages exceed those reported for experts, who varied their contributions by more often including questions and meta-statements.)

Students from both conditions demonstrated significant improvement from pretest to posttest with regard to the frequencies of Counter-C units, $z = -3.66, p < .001$ for the experimental condition and $z = -2.43, p = 0.02$ for the comparison condition. They were also more likely to have Counter-C units in their constructed dialogs at posttest than at pretest, $\chi^2(1, 32) = 10.71, p < .001$ for the experimental condition and $\chi^2(1, 40) = 4.76, p = 0.03$ for the comparison condition.

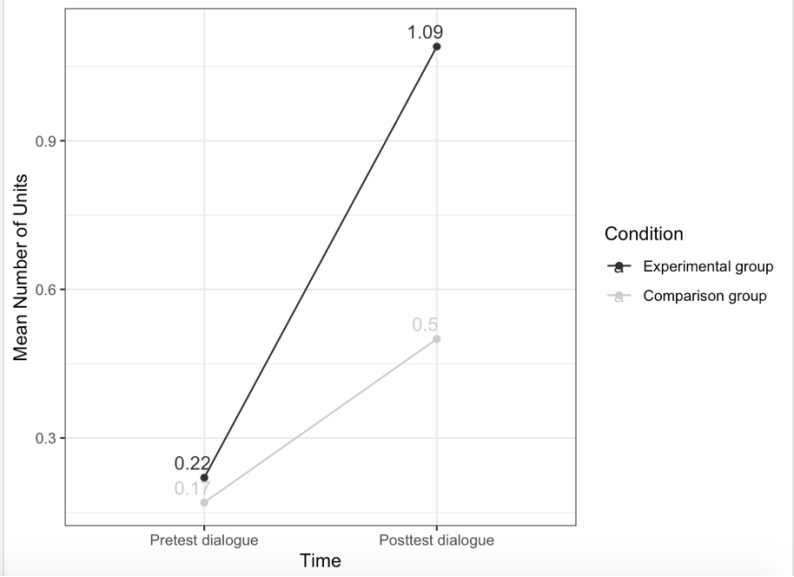


Figure 5. Counter-C Frequencies in Constructed Dialogs by Condition and Time.

Counter-U frequencies. The mean occurrences of idea units in the Counter-U category at pretest and posttest appear in Figure 6. As shown, at pretest, experimental students showed 0.19 Counter-U unit (SD = 0.47) while comparison students had no such units at all, a nonsignificant condition difference ($\hat{\beta} = 19.63, p = 1$). At posttest, the mean number of Counter-U units in constructed dialogs was 1.0 (SD = 1.37) in the experimental condition and 0.17 (SD = 0.38) in the comparison condition. According to Table 6, after controlling for pretest performance, at posttest, experimental students had a notable 6.40 times more counter-U units in their constructed dialogs than did comparison students, a significant advantage ($\hat{\beta} = 1.86, p < .001$).

Who contributed to such occurrences? At pretest, the percentage of students who generated at least one Counter-U unit was 16% for the experimental condition and 0 for the comparison condition, a significant difference, $p = 0.01$ ($Fisher's\ Exact\ Test$). At posttest, 53% of students from the experimental and 18% from the comparison condition did so, again a significant difference, $\chi^2(1, 72) = 10.15,\ p = .001$.

Significant improvement from pretest to posttest occurred among students from both conditions in the frequencies of Counter-U units in their constructed dialogs, $z = -2.79, p = 0.01$ for the experimental condition and $z = -2.55, p = 0.01$ for the comparison condition. Students from both conditions also became more likely to generate a Counter-U unit in their posttest dialogs compared to pretest, $\chi^2(1, 32) = 7.20, p < .001$ for experimental condition and $p = 0.02$ $(Exact\ Binomial\ Test)$ for comparison condition.

The significant within-condition progressions in mastering and applying the Counter-C and Counter-U strategies attest to the general efficacy of the intervention in developing students' strong counterargument strategies, as well as to the added contribution of expert arguers to the progress made by students in the experimental condition.

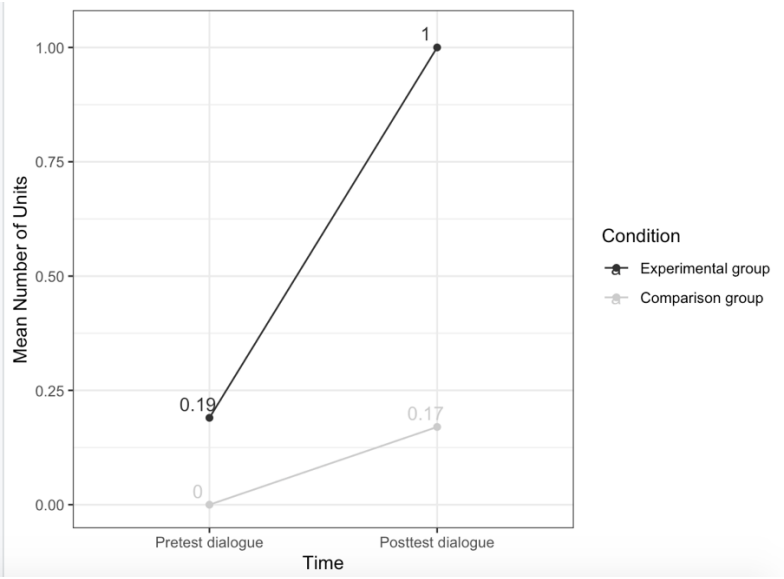


Figure 6. Counter-U Frequencies in Constructed Dialogs by Condition and Time.

Concession frequencies. The mean numbers of idea units classified into the concession category at pretest and posttest appear in Figure 7. As shown, at pretest, experimental students on average showed 0.22 (SD = 0.55) concession units in their constructed dialogs and comparison

students showed a similar mean number (Means = 0.1, SD = 0.30); the condition difference at pretest was not significant ($\hat{\beta} = 0.81, p = 0.24, IRR = 2.24$). At posttest, while the mean number of Concession units in experimental students' dialogs was 0.69 units (SD = 0.82), the mean number in comparison students' dialogs remained 0.1 units (SD = 0.30). According to Table 6, after controlling for pretest performance, at posttest, experimental students had a notable 6.34 times more Concession unit in their constructed dialogs than comparison students, a significant advantage ($\hat{\beta} = 1.86, p < .001$).

Who was responsible for the occurrences of concession units? We observed that at pretest, the percentage of students who made at least one concession unit was 16% in the experimental group and 10% in the comparison group, a nonsignificant difference, $p = 0.49$ ($Fisher's\ Exact\ Test$). At posttest, this percentage was 50% for the experimental group and remained 10% for the comparison group, a significant condition difference, $\chi^2(1, 72) = 14.18,\ p < .001$.

Within the experimental condition, significant improvement occurred from pretest to posttest in the frequencies of concession units in constructed dialogs, $z = -2.81, p = 0.05$, as well as the number of students contributing to these occurrences, $\chi^2(1, 32) = 9.31, p < .001$. No change in these respects was observed on the part of the comparison students.

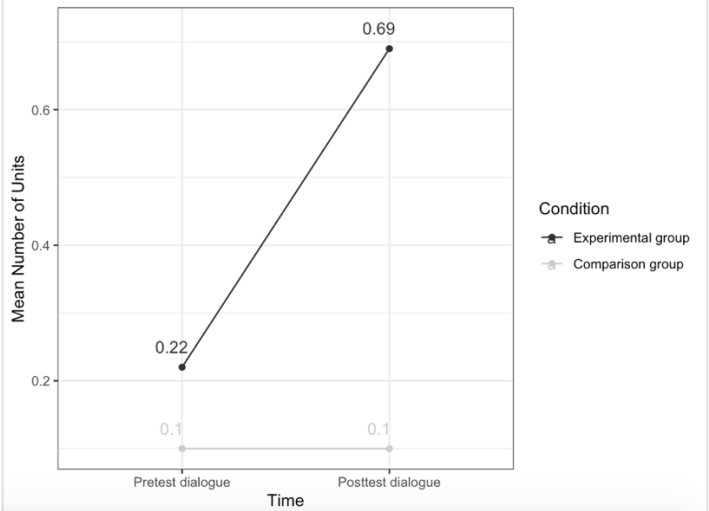


Figure 7. Concession Frequencies in Constructed Dialogs by Condition and Time.

Unconnected frequencies. The mean frequencies of the unconnected unit, at pretest and posttest are shown in Figure 8. At pretest, experimental students on average showed 0.97 (SD = 1.28) unconnected units in their constructed dialogs in comparison to 0.49 (SD = 0.78) such units in dialogs constructed by comparison students, a significant difference ($\hat{\beta} = 0.69, p = 0.05, IRR = 1.99$). At posttest, the mean number of Unconnected units decreased to 0.5 (SD = 0.72) units in experimental students' dialogs, while increasing to 0.82 (SD = 1.55) units in comparison students' dialogs. According to Table 6, after pretest performances were controlled for, at posttest experimental students showed half as many unconnected units as comparison students, a non-significant difference ($\hat{\beta} = -0.59, p = 0.14, IRR = 0.55$). The increased number of idea units reported in experimental students' post-intervention dialogs thus were not due to an increase in unconnected utterances.

At pretest, the proportion of students who showed at least one Unconnected unit was 50% and 34% for experimental and comparison students, respectively, a nonsignificant difference, $\chi^2(1, 73) = 0.06,\ p = 0.81$. At posttest, these percentages were 38% and 45% for experimental

and comparison conditions, respectively, again a nonsignificant difference, $\chi^2(1, 72) = 0.41$, $p = 0.52$.

In the experimental condition, the decrease from pretest to posttest in the mean frequencies of unconnected units was nonsignificant, $z = -1.84, p = 0.07$ as it was in the number of contributing students, $\chi^2(1, 32) = 1.14, p = 0.29$. There were no significant changes in these respects over time on the part of the comparison students, in either frequencies of occurrences, $z = -1.09, p = 0.28$, or number of contributing students, $\chi^2(1, 40) = 1.27, p = 0.26$.

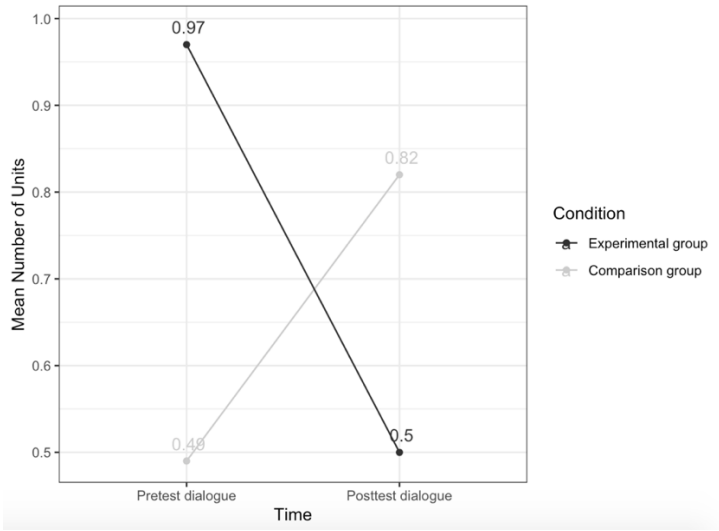


Figure 8. Unconnected Frequencies in Constructed Dialogs by Condition and Time.

Evidence-based frequencies. The mean numbers of idea units in the Functional Evidence-based category appear in Figure 9. As shown, at pretest, experimental and comparison students showed comparable numbers of idea units successfully justified by evidence (means = 0.56 and 0.54, SDs = 0.98 and 0.95, for experimental and comparison groups, respectively); the slight condition difference at pretest was not significant ($\hat{\beta} = 0.05, p = 0.91, IRR = 1.05$). At posttest, while the mean frequency in experimental students' constructed dialogs increased to

1.25 units (SD = 1.39), the frequency decreased to 0.48 (SD = 0.68) units in comparison students' dialogs. As seen in Table 7, after controlling for pretest performance, experimental students showed an expected 2.63 times more evidence-based units than did comparison students, a significant difference ($\hat{\beta} = 0.97, p < .001$).

At pretest, 34% and 32% of students from the experimental and comparison conditions made at least one use of evidence, respectively, a nonsignificant difference, $\chi^2(1, 73) = 0.06$, $p = 0.81$. At posttest, 63% of the experimental group did so, versus 38% from the comparison group, a significant difference, $\chi^2(1, 72) = 4.45,\ p = 0.03$.

Within the experimental condition, significant increases from pretest to posttest occurred in both frequencies of Evidence-based units, $z = -2.80, p = 0.01$, and proportion of students who contributed to such occurrences, $\chi^2(1, 32) = 6.23, p = 0.01$. No changes in these respects reached significance on the part of the comparison students, either in frequencies of occurrence, $z = -0.46, p = 0.65$, or proportion of contributing students, $\chi^2(1, 40) = 0.40, p = 0.53$.

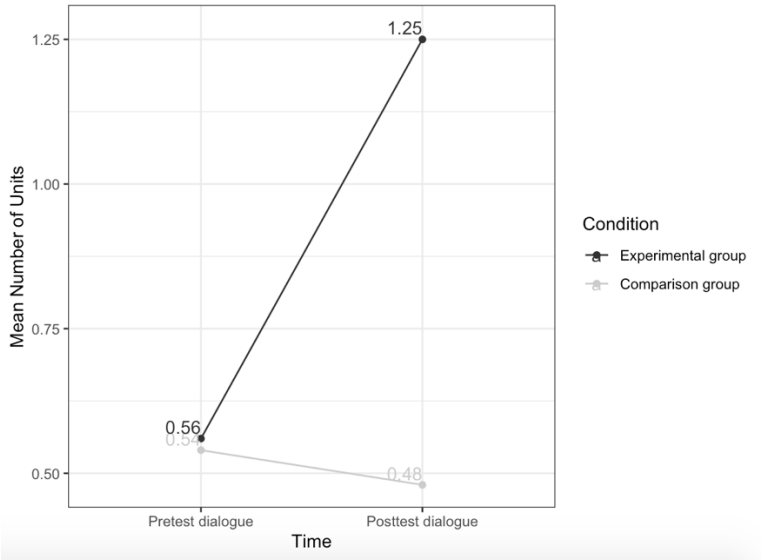


Figure 9. Evidence-based Idea Unit Frequencies in Constructed Dialogs by Condition and Time.

Table 7. Estimation Results of the Negative Binomial Regressions on Posttest Constructed Dialogs: Evidence-based Types.

Response Variable	$\hat{\beta}$	Estimates	Std Err of $\hat{\beta}$	Conf. Int. of $\hat{\beta}$	IRR[c]	Sig.
Functional	Intercept	-1.07	0.26	(-1.62 to -0.58)		<.001
Evidence-based	Condition[a]	0.97	0.29	(0.41 to 1.56)	2.63	**<.001**
Idea Units	Pretest[b]	0.41	0.11	(0.19 to 0.62)	1.51	<.001

[a] condition = 0 which was the peer-only group that was taken as the reference group.
[b] pretest performance on various outcome variables were taken as covariate in respective models.
[c] the incidence rate ratios.

Successful *However*-pair frequencies. Figure 10 displays the mean number of *However* pairs in students' constructed dialogs. At pretest, both experimental and comparison students showed little to no competence in this respect (Means = 0.12 and 0, SDs = 0.34 and 0 for experimental and comparison students, respectively); the slight condition difference in mean frequencies was not significant ($\hat{\beta} = 20.22, p = 1$). At posttest, the mean number of *However* pairs increased to 0.44 (SD = 0.72) in the experimental condition and to 0.07 in the comparison condition (SD = 0.27). As seen in Table 8, after controlling for pretest performance, experimental students at posttest had an expected and notable 5.24 times more *However* pairs than did comparison students, a significant difference ($\hat{\beta} = 1.66, p = 0.01$).

At pretest, the percentage of students contributing to occurrences of *However* pairs was 13% and 0% for the experimental and comparison conditions, respectively, a significant difference, $p = 0.03$ ($Fisher's\ Exact\ Test$). At posttest, this number was 34% and 8% for the experimental and comparison conditions, respectively, again a significant difference, $\chi^2(1, 72) = 6.57,\ p = 0.01$.

Within the experimental condition, significant increases from pretest to posttest occurred in frequencies of *However* pairs, $z = -2.25, p = 0.02$, as well as number of students who contributed to such occurrences, $\chi^2(1, 32) = 5.44, p = 0.02$. No parallel, significant

improvement in these respects was observed on the part of the comparison students, for frequencies of occurrences, $z = -1.44, p = 0.15$, or for the percentage of contributing students, $p = 0.25$ $(exact\ Binomial\ Test)$.

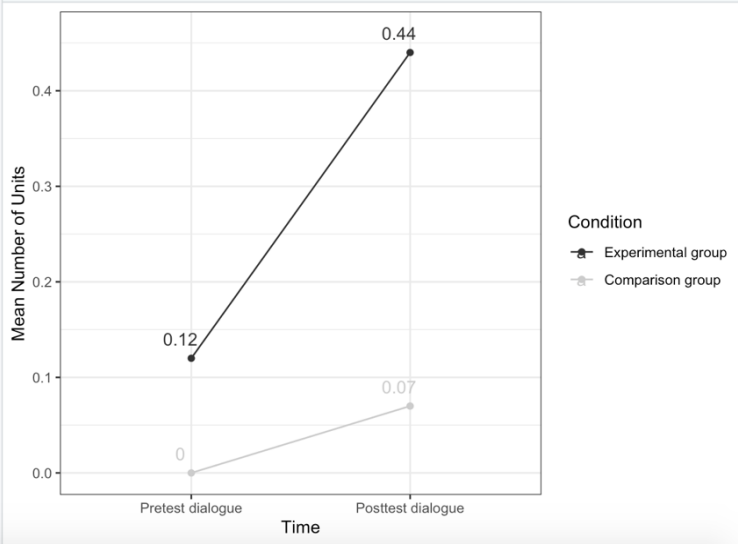


Figure 10. Successful *However* Compound Frequencies in Constructed Dialogs by Condition and Time.

Table 8. Estimation Results of the Negative Binomial Regression on Posttest Constructed Dialogs: However Pairs.

Response Variable	$\hat{\beta}$	Estimates	Std Err of $\hat{\beta}$	Conf. Int. of $\hat{\beta}$	IRR[c]	Sig.
Successful	Intercept	-2.59	0.58	(-3.99 to -1.63)		<.001
However	Condition[a]	1.66	0.66	(0.47 to 3.15)	5.24	**.01**
Compound Unit	Pretest[b]	0.65	0.69	(-0.89 to 1.89)	1.91	.35

[a] condition = 0 which was the peer-only group that was taken as the reference group.
[b] pretest performance on various outcome variables were taken as covariate in respective models.
[c] the incidence rate ratios.

Summary of experimental group advances across time. A summary of intervention effects on experimental-group outcome skill indicators appears in Table 9 for reference in showing how widely exercised (both across and within groups) each skill indicator had become. In the comparison group, time effects were also often significant, as reported.

Table 9. Summary of Intervention Effects on Experimental-Condition Outcome Skill Indicators: Constructed Dialog Task.

Indicator	Effects	% showing		Mean usage	
		Initial	Final	Initial	Final
On-topic Idea units	Condition, Time	100	100	3.47	5.72
Question	Time	9	28	0.09	0.34
Meta-talk	Time	3	19	0.03	0.19
Indirect Counters	Time	25	47	0.28	0.59
Counter-Critique	Condition, Time	22	69	0.22	1.09
Counter-Undermine	Condition, Time	16	53	0.19	1.00
Concession	Condition, Time	16	50	0.22	0.69
Unconnected Unit	Condition	50	38	0.97	0.5
Evidence-based Idea Unit	Condition, Time	34	63	0.56	1.25
Successful *However* pair	Condition, Time	13	34	0.12	0.44

3.4 Statistical Analysis of Essays

The main statistical analysis of post-intervention essays on a new topic again consisted of negative binomial analysis, supplemented by analysis of individual patterns of usage.

The number of essays analyzed was 31 for the experimental and 39 for the comparison conditions; four students didn't write the essay test due to absence.

Idea Units. As shown in Figure 11, the post-intervention essays written by experimental and comparison students contained similar numbers of on-topic idea units Means = 8.58 and 8.49, SDs = 3.94 and 2.65, for experimental and comparison groups, respectively, a nonsignificant difference, $\hat{\beta} = 0.01, p = 0.90, IRR = 1.01$(Table 10).

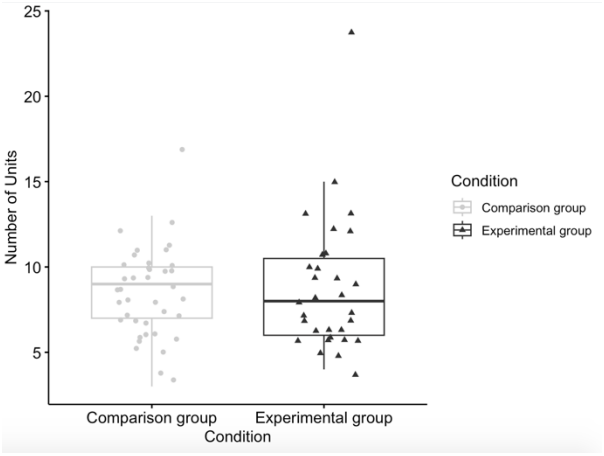


Figure 11. On-topic Idea Units in Transfer Essay by Condition.

Table 10. Estimation Result of the Negative Binomial Regression on Transfer Essay: Idea Units.

$\hat{\beta}$	Estimates	Std Err of $\hat{\beta}$	Conf. Int. of $\hat{\beta}$	IRR[b]	Sig.
Intercept	2.14	0.06	(-0.11 to 0.13)		<.001
Condition[a]	0.01	0.09	(-0.16 to 0.19)	1.01	.90

[a] condition = 0 which was the peer-only, comparison group that was taken as the reference group
[b] the incidence rate ratios

Support-own (M+) frequencies. As shown in Figure 12, experimental students' essays contained more idea units functioned to support their own position than did comparison students' (Means = 3.68 and 2.79, SDs = 1.47 and 1.70 for experimental and comparison groups, respectively). Negative binomial regression analysis (Table 11) showed experimental students had 1.32 times more M+ units than comparison students in their essays, a significant difference ($\hat{\beta} = 0.27, p = 0.04$). All students in the experimental condition and 92% in the comparison condition included M+ idea units in their essays, a nonsignificant difference ($Fisher's\ Exact\ Test$, p = 0.25).

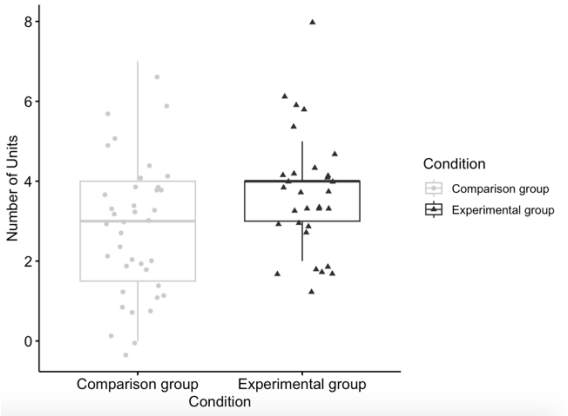


Figure 12. M+ Idea Units in Transfer Essay by Condition.

Table 11. Estimation Results of the Negative Binomial Regressions for Transfer Essay.

Response Variable	$\hat{\beta}$	Estimates	Std Err of $\hat{\beta}$	Conf. Int. of $\hat{\beta}$	IRR[b]	Sig.
Support-own	Intercept	1.03	0.10	(0.83 to 1.21)		<.001
	Condition[a]	0.27	0.13	(0.01 to 0.54)	1.32	**.04**
Weaken-other	Intercept	0.61	0.13	(0.35 to 0.87)		<.001
	Condition	0.60	0.18	(0.25 to 0.95)	1.82	**<.001**

[a] condition = 0 which was the peer-only group that was taken as the reference group
[b] the incidence rate ratios

Weaken-other (O-) frequencies. As shown in Figure 13, experimental students' essays contained more idea units that functioned to weaken the opposing side's claim than did those of comparison students' (Means = 3.35 and 1.85, SDs = 2.11 and 1.67). Experimental students had an expected 1.82 times more O- idea units in their essays than did comparison students, a significant difference (Table 11; $\hat{\beta} = 0.60, p < .001$). All students in the experimental condition and 72% in the comparison condition included one or more O- units in their essays, a significant difference (Fisher's Exact Test, $p < .001$).

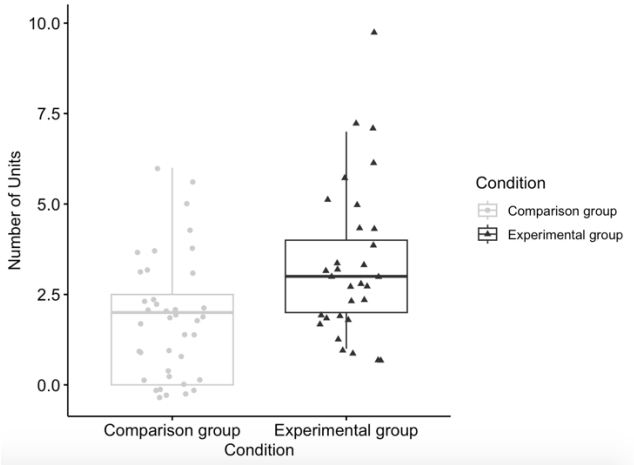


Figure 13. O- Idea Units in Transfer Essay by Condition.

Belief-incongruent frequencies. Addressing belief-incongruent arguments is a bigger challenge for students than building their own-side arguments or seeking to weaken other-side arguments, and many do not attempt to do so. Incidences of belief-incongruent idea units in transfer essays appear in Figure 14. Means = 1.0 and 0.54, SDs = 1.06 and 0.79, for experimental and comparison students, respectively. Experimental students showed an expected 1.86 times more belief-incongruent idea units than did comparison students, a significant advantage (Table 12; $\hat{\beta} = 0.62, p = 0.04$). About two thirds (61%) of experimental students versus slightly more than one third (38%) of comparison students contributed to the occurrences of belief-incongruent units, a difference that just fails to reach significance, $\chi^2(1, 70) = 3.60, p = 0.06$.

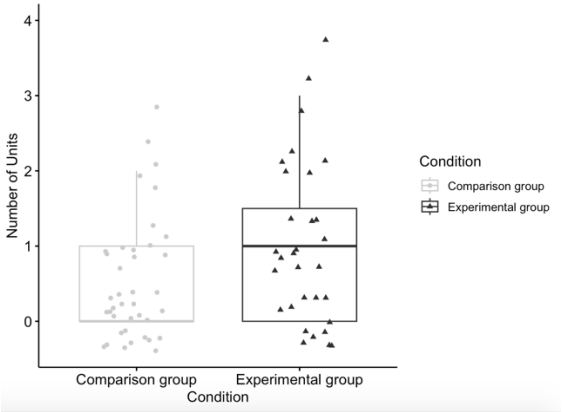


Figure 14. Belief-incongruent Idea Units in Transfer Essay by Condition.

Table 12. Estimation Results of the Negative Binomial Regressions for Transfer Essay: Belief-incongruent type.

Response Variable	$\hat{\beta}$	Estimates	Std Err of $\hat{\beta}$	Conf. Int. of $\hat{\beta}$	IRR[b]	Sig.
Belief-incongruent	Intercept	-0.62	0.23	(-1.09 to -0.20)		0.01
(O+ or M-)	Condition[a]	0.62	0.30	(0.04 to 1.22)	1.86	**0.04**

[a] condition = 0 which was the peer-only group that was taken as the reference group
[b] the incidence rate ratios

Evidence-based frequencies. To what extent were arguments supported by information that functioned as evidence? As shown in Figure 15, experimental students' essays contained more evidence-based units than did those of comparison students. Means = 4.29 and 2.56, SDs = 2 and 2.17, respectively. Experimental students had an expected 1.67 times more evidence-based idea units than comparison students, a significant difference (Table 13; $\hat{\beta} = 0.51, p < .001$). All experimental students versus 87% of comparison students included at least one evidence-based unit, a difference just failing to reach significance (Fisher's Exact Test, $p = 0.06$).

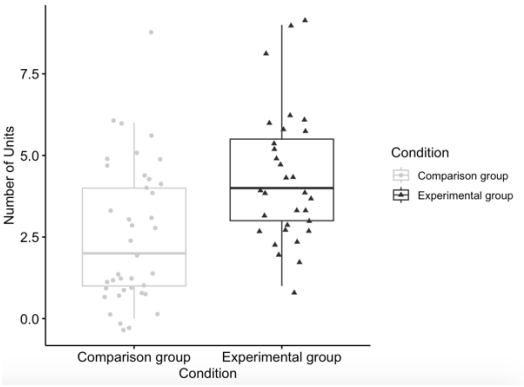


Figure 15. Evidence-based Idea Units in Transfer Essay by Condition.

Table 13. Estimation Results of the Negative Binomial Regression on Transfer Essay: Evidence-based Idea Units.

Response Variable	$\hat{\beta}$	Estimates	Std Err of $\hat{\beta}$	Conf. Int. of $\hat{\beta}$	IRR[b]	Sig.
Functional Evidence-based	Intercept	0.94	0.11	(0.72 to 1.16)		<.001
	Condition[a]	0.51	0.15	(0.22 to 0.81)	1.67	**<.001**

[a] condition = 0 which was the peer-only group that was taken as the reference group
[b] the incidence rate ratios

***However*-pair frequencies.** Means for successful *However* pairs in students' essays were 0.71 and 0.26 (SDs = 0.94 and 0.50) for experimental and comparison students, respectively. Experimental students produced 2.77 times more successful *However* unit pairs than did comparison students, a significant difference, $\hat{\beta} = 1.02, p = 0.01$. A total of 45% of experimental students generated at least one such unit, compared to 23% of comparison students, a significant difference, $\chi^2(1, 70) = 3.82, p = 0.05$. See Figure 16 and Table 14.

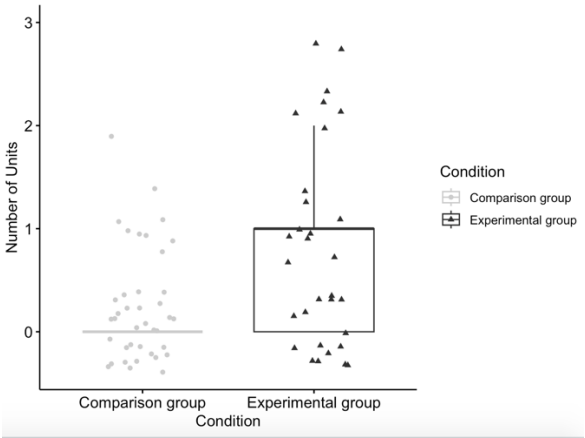


Figure 16. Successful However Compound Unit in Transfer Essay by Condition.

Table 14. Estimation Results of the Negative Binomial Regression on Transfer Essay: However Unit-pair.

Response Variable	$\hat{\beta}$	Estimates	Std Err of $\hat{\beta}$	Conf. Int. of $\hat{\beta}$	IRR[b]	Sig.
Successful However Compound Unit	Intercept	-1.36	0.33	(-2.07 to -0.77)		<.001
	Condition[a]	1.02	0.40	(0.26 to 1.84)	2.77	**0.01**

[a] condition = 0 which was the peer-only group that was taken as the reference group
[b] the incidence rate ratios

Summary of Intervention Effects Across Conditions. A summary of intervention effects on outcome skill indicators for both conditions appears in Table 15.

Table 15. Summary of Intervention Effects on Outcome Skill Indicators: Transfer Essay Task.

Indicator	% showing		Mean usage	
	Comparison	Experimental	Comparison	Experimental
On-topic Idea units	100	100	8.49	8.58
Support-own	92	100	2.79	3.68*
Weaken-other	72	100*	1.85	3.35*
Belief-incongruent	38	61	0.54	1.0*
Evidence-based	87	100	2.56	4.29*
However pair	23	45*	0.26	0.71*

*significant condition difference, $p <= .05$ (exact p-values are reported in respective results sections above).

Chapter 4: Discussion

Extravagant claims are often made for the power of discourse. If only those involved talk to one another, we hear said, problems small and large become resolvable. In fact, successful resolution is not the most common outcome, and success if it comes at all will come only with the sustained motivation and skill of its partakers. It is known that wide individual differences exist in argument skills and affect outcomes, as the present results illustrate. Yet we know less than we need to about the mechanisms of development of these skills, given their importance in very many spheres of individuals' lives.

In the present investigation of argument skills, we adopt an apprenticeship model of intellectual development that has been widely endorsed with respect to both cognitive and behavioral skills emerging in the first decade of life, in particular literacy skills. Skills advance through participation in activities that afford practice in these skills in the company of more skilled individuals who both guide and serve as role models (Bridgers et al. 2023; Harris, 2012). But apprenticeship with more skilled role models has not received such attention in the case of higher-order cognitive skills such as argumentation that continue to develop into and beyond the second decade.

It is to fill this gap that motivated our implementation of the present study. Our results in general confirmed our hypotheses in showing that the experimental students who had the opportunity to argue with a series of older and more capable expert arguers advanced more in argument skills, relative to their counterparts in the comparison condition who engaged in the same intervention program except the benefit of discourse with a more skilled partner. In addition, since dialogic argumentation is the proposed mechanism of advancement (Hemberger et al., 2017; Kuhn et al., 2016; Shi et al., 2019), we also established that gains occurred in

dialogic argumentation skill as well as conventional argument writing. An advantage of the present study is that by adopting a range of outcome measures, we were able to pinpoint more precisely where the gains occurred and where they did not. These insights carry both theoretical and educational implications. In what follows, I first summarize the principal findings of the present study, followed by discussion of its theoretical and educational implications. I will end this chapter by pointing out the limitations of the current study and possible future directions.

4.1 Summary of Results

Expert Modeling of Advanced Dialogic Contributions

As evaluations of our experimental manipulations, we first demonstrated that participants, blind to the identities of their discourse partners, perceived the dialogic contributions of their expert partners as more skilled than those of peer partners. Then, qualitative examination of experts' argumentation revealed that in dialogs they often modeled use of evidence to support claims and skilled forms of counterarguments, as well as frequent questioning with respect to the partner's statements and meta-talk about the discourse itself (e.g., "Do we agree on this?" or "Is this what you're saying?"). These effective strategies were appropriated and later implemented individually by the younger participants to varying degrees in their writings.

Effects on Dialogic Argumentation Skills on a Transfer Topic

Results derived from the analyses of students' solitarily constructed dialogs in general supported the apprenticeship model as a developmental mechanism of students' dialogic argumentative skills. Relative to those who did not have the benefit of discourse with a more skilled partner, the experimental students showed superior performance in constructing skilled

counterarguments and in coordinating arguments with evidence bearing on them, skills deemed central to productive and constructive argumentation (Kuhn, 1991; Macagno, 2016). Furthermore, following their more expert opponents, they showed advancement of integrative skills, making not only more Concession statements acknowledging the merits of an opposing argument, but also more However unit-pairs coordinating opposing claims while seeking to advance their own arguments at posttest, in comparison to their pretest performance. In contrast, comparison students showed little development in making integrative arguments. Additionally, experimental students also made less Unconnected statements at posttest relative to comparison students, suggesting that they had a better understanding of the purposes of argumentation.

However, while in dialogs a considerable proportion (more than 50%) of their expert partners' contributions were probing units comprising of either questioning statements regarding the content, validity, and/or understanding of an argument or meta-talk statements about the dialogic processes, neither strategy was adopted to a significant degree by the younger participants when they were reasoning alone. This, we suspect, was caused by an inadequate meta-strategic understanding of these argumentive strategies which we will elaborate more later in the theoretical implications section.

The finding that students in the comparison condition, who experienced the same dialogic argumentation intervention except the opportunity of conversing with experts, also advanced from pretest to posttest in terms of skilled counterarguments aligns with previous research in showing the efficacy of the AWM program in developing students' dialogic argumentation skills (see Iordanou & Rapanta, 2021 for a review). However, the fact that they showed little growth in skills of theory-evidence coordination and integrative argumentation also echoes with previous findings in suggesting the necessity of additional scaffolding to promote and accelerate the

development of more complex and challenging skills of argument (Hemberger et al., 2017; Kuhn & Moore, 2015; Shi, 2019).

Effects on Individual Argumentive Writing on a Transfer Topic

Results derived from the analyses of students' individually composed argumentative essays on a transfer topic lent strong support for the benefits of the apprenticeship model in developing individual argument skills. Students in the experimental group outperformed their counterparts in the comparison group in all the four skill indicators that are suggestive of advanced abilities, namely the frequencies of weaken-other, belief-incongruent, evidence-based, and However statements in their essays.

Although not directly comparable, juxtaposing students' posttest dialog and essay performance might reveal nuances regarding the affordances of the two assessment tools as well as students' understandings of the purposes of the two genres of writing (Shi, 2020a). Specifically, consistent with previous studies (Felton et al., 2015; Hemberger et al., 2017; Shi, 2019, 2020a), we observed that in the present study students in both conditions demonstrated obvious myside-bias in essays since they devoted disproportionate number of statements to support their favored position (either by supporting it directly or indirectly by weakening the opposing position). On the contrary, in dialogs, they appeared to be more capable of addressing the strengths and weaknesses of alternative positions by constructing counterarguments while adopting and alternating between the two positions (Shi, 2020a). However, in terms of the two more demanding skills, namely the coordination of arguments/relevant evidence and the reconciliation of opposing arguments, students appeared to display greater competence in essays than in dialogs, even though the latter has the presumed benefit of an (imagined) interlocutor. Anecdotal evidence from interviews with students following their completion of the

interventions and assessments suggested that students in general perceived the dialog task as more challenging than the essay task. They indicated that the dialog task was more demanding and distracting since they had to reason about, evaluate, counter, and defend two positions simultaneously. In contrast, the essay task was more familiar and easier to them so that they were able to think deeper about their favored positions.

4.2 Theoretical Implications

The apprenticeship model of argumentive skill development

The present work advances a model of apprenticeship as the mechanism in the development of argument skills in an important way. Specifically, it is to our knowledge the first demonstration of skill advancement in a traditional measure of argument skill – a brief individual essay – following engagement and practice in dialogic argumentation with both more and similarly skilled others, in the absence of direct writing or other instruction. Since dialog is the proposed mechanism of advancement, we establish, moreover, that these gains occur in dialogic argumentation skill as well as conventional argument writing. The experimental design allowed us to rule out several alternative explanations for the experimental group's superior performance. First, the inclusion of a carefully controlled comparison group who differed with the experimental group only in terms of their electronic discourse partner group composition enabled us to eliminate confounding factors caused by variations in instructions (e.g., the timing, amount, and types of instructions). In addition, since the text-only medium allowed substitution of the more skilled dialogic partner in the absence of the awareness of the target participant, we were able to remove confounding social factors as well. Furthermore, the evidence students had access to at posttest dialog and essay assessments was in text format, following the standard practice at

schools. In contrast, presenting evidence in the Q&A format, which occurred during the intervention, had been found to facilitate students' use of evidence in arguments by signaling and exemplifying its potential functions (Iordanou, et al., 2019; Shi, 2019). By removing the benefits of the Q&A format and presenting information in its raw form, we were able to show that any advancements students demonstrated in coordinating arguments and relevant evidence were due to their own improvement at the competence as well as dispositional levels, rather than cued by the task materials (Fedyke et al., 2019; Kuhn, 2022; Kuhn & Zillmer, 2015; Kuhn et al., 2013).

Regarding the debate between Vygotskian and Piagetian conceptualizations of the mechanisms of intellectual development, specifically the relative merits of adult-child versus peer-peer discourses, our results lent support for the former viewpoint. Where do the benefits of a more competent dialogic partner lie in the development of argument skills? Sfard (2015), on the basis of conceptualizing learning as "gaining mastery over a well-defined form of communication" (p250), made a distinction between object-level and meta-level changes of discourse (i.e., learning). She claimed that the latter depended on adult-child (or teacher-student) discourses while the former could happen through peer-peer discussions. The reason is that in contrast to object-level discourse which aims to extend the content of a discourse, meta-level discourse development zeroes in on changing the nature or norms of a discourse within certain social, cultural, and historical contexts. Those with less experience and/or knowledge lack the intuitions and metacognitive competence (Kuhn, 2022) to make the appropriate decisions regarding the adoption of one form of practice (and correspondingly, the abandonment of the other). Thus, the demonstration of more advanced practices and the social regulation provided by an expert becomes the indispensable first step in initiating a cascade of developmental changes in a novice within her ZPD. The co-participation of the novice who observes the effectiveness of

new strategies and whose behaviors are socially scaffolded/regulated by the expert is the necessary next step, ideally followed by the novice gradually taking over the task of (self-) regulating her own actions (i.e., the transfer of scaffolding from external to internal sources) and eventually gaining full ownership of the practice which become automatized over time with exercises (Shvarts & Bakker, 2019). In contrast, once children learn about the rules of a certain type of dialog, they can use these rules to generate new narratives through independent or collaborative peer dialogs (Kuhn, 2015).

The electronic dialog sessions constituted powerful contexts for the potentials of expert modeling and scaffolding to actualize. One the one hand, the dialog contexts are characterized by strong and direct social engagement and (thus) demand in the sense that their participants are expected, motivated, and socially regulated to respond relevantly to one another's statements. In argumentative dialogs, the relevance of an argument is determined by its pragmatic function in relation to the goal of the dialog, which can be succinctly summarized as to increase (or decrease) the acceptability of a position (Macagno, 2016). When an arguer's argument is undermined by an opposing partner, she has to restructure her reasoning in order to restore her position based on the received "feedback". In the meantime, she is motivated to impose an at least equally powerful "attack" on the opponents' arguments. In the case where the opponent purposefully modelled more advanced strategies, the less skilled arguer benefits by accommodating her behaviors toward those demonstrated by the expert. On the other hand, the electronic medium has the potential to facilitate participants' reflections on the ongoing dialog by providing a written record of it. This contrasts with oral argumentation in which statements disappear as soon as spoken. Previous research has shown that students' growth in metacognitive competence and dispositions are critical byproducts of the AWM approach that arguably drives

sustained engagement and development at the strategic level and transfer of skills to novel contexts or even domains (Iordanou, 2010; Kuhn, 2019, 2022; Kuhn et al., 2011, 2013; Shi, 2019).

As support to the above theoretical speculations, we observed that during the electronic dialog sessions, experts frequently probed and (thus) scaffolded the younger participants' reasoning and arguments by devoting more than half of their dialogic contributions to either question the acceptability/tenability of their arguments or to (metacognitively) regulate the argumentation processes. Once the younger participants endorsed the standards set by the experts by observing their effectiveness in advancing and deepening the dialogs, they not only accommodated their own behaviors to match the new standards but also imposed the same requirements for their opponents' arguments. This implicates growth at the meta-strategic level regarding the purposes of arguments and various strategies to achieve them (Kuhn & Zillmer, 2015; Kuhn et al., 2013) on the part of the younger novice arguers. In our post-intervention assessments of argument skills in both dialogs and essays, the results showed that the experimental group, following their more skilled partners, displayed superior performance than the comparison group in terms of the generation of advanced counterarguments, the coordination of arguments with relevant evidence, and the integration and reconciliation of conflicting arguments (i.e., However pairs). All of these strategies presuppose the applications of metacognitive skills resulting in adequate representations, deeper evaluations, and integrative coordination of the merits and weaknesses of alternative arguments and evidence (Mayweg-Paus et al., 2015; Kuhn, 2022; Shi, 2019, 2020a). On the contrary, the characteristic of the discourse partner seemed to have little impact on the lengths of essays students wrote as well as their

ability to propose additional arguments (e.g., indirect counters) which had little metacognitive demand.

The transfer of skills from dialogic to individual written argument

Another theoretical implication of the present study is its demonstration of the transfer of skills from the social dialogic to the individual planes, the so-called sociocultural learning posited by Vygotsky (1978, 1984) and his successors (Rogoff, 1990, 1995). Previous research has demonstrated that sustained engagement in dialogs bridges skill gaps between the inter-psychological and the intra-psychological planes (Hemberger ger et al., 2017; Kuhn et al., 2016; Resnick et al., 2015), and the present study substantiated this notion by showing that skill gains occurred in individual dialogic argumentation as well as conventional argument writing, following engagement in intense and extended argumentation practices with others. Importantly, building on previous studies (Halpern, 2022; Mayweg-Paus et al., 2015; Papathomas & Kuhn, 2017), our research design enabled us to address two competing alternative explanations for the alleged effects of dialogs. On the one hand, being mindful that the characteristic (e.g., their knowledge and skill levels) of the interlocutor might interfere with the target participants' dialogic performance (Halpern, 2022), in the present study, we removed this confounding factor by having participants independently construct an argumentative dialog between two skilled arguers discussing a meaningful social issue. Thus, the qualities of the constructed dialogs represented the participants' unimpacted genuine understandings of and skills to engage in argumentation, an intricately interactive practice. They were also cued to display their best performance since the dialog was framed to happen between two expert arguers.

On the other hand, since the two post-intervention assessments were both framed within a novel unstudied topic, they constitute transfer tasks (but of varying distance from the

intervention activities). Thus, we were able to reject the explanation that what we observed was (at least partially) participants' memory capabilities to re-produce the artifacts of socially shared cognition co-constructed by opposing interlocutors during the electronic dialog sessions. Rather, the experience of argumentation did engender something qualitatively new in the way they reason about controversial issues that altered their approaches to new tasks even in the absence of a dialogic partner.

Then, what changed in students' minds through their dialogic experience that were later transferred to individual reasoning tasks? Here we once again invoke the role played by metacognition, especially its motivational aspects (Kuhn, 2019, 2022). We suspect both groups experience some levels of metacognitive development. As discussed above, metacognitive growth enabled the experimental group to adopt more sophisticated and pragmatically relevant argumentive strategies modelled by the experts. In the meantime, we also observed that compared with their pre-intervention performance, both groups advanced in skilled counterarguments in dialogs as well as showed remarkable performance in envisioning and weakening a potential competing position in essays. Both areas of advancement implicate the applications of strategic metacognitive skills of argumentation (Iordanou & Rapanta, 2021). This also suggested that both groups interiorized a dialogic framework for reasoning following their experience of argumentation (Kuhn, 2019). The dialogic framework provides the "missing interlocutor" who gives individual discourse an audience and a purpose to its exchanges. Experiencing a flesh-and-blood opposing partner during the electronic dialog sessions enabled participants to abstract and then interiorize this dialogic framework and when they were later asked to express themselves individually in writing, an "oppositional voice" (Ford & Forman,

2012) appeared and motivated them to structure their arguments in response to what others might say.

It has been reported that during the electronic dialog sessions, peers of comparable skill levels took on the role of meta-level support provider and receiver interchangeably and as needed, flexibly scaffolding one another's development and understandings until skills gain stability (Zillmer & Kuhn, 2018). But the fact that the experimental group demonstrated superior and wider gains (e.g., improvement in integrative arguments) than the comparison group echoes with previous studies in showing that metacognitive development happened at different paces depending on the levels of scaffolding (Shi, 2019, 2020b).

The observation that students from neither condition showed satisfactory usages of meta-talks and questioning statements in their individually constructed dialogs might raise doubts about our attribution of positive transfer effects to metacognitive gains. But two things need to be considered before we make any final judgement. One is that previous research using the same assessment format has shown that constructed dialog task was conducive to stimulate participants' explorations of the problem space (Zavala & Kuhn, 2017) and uncover their abilities to represent and address the weaknesses of alternative views (Shi, 2020a). In our study, we did observe students devoted a large (roughly 50% in the experimental condition) proportion of their writings to counterargument statements. A focus on discussing the content of the topic left them with little space to talk about the processes of argumentation, an inclination that fits well with the school-related genre of writing (Kuhn & Moore, 2015). Relatedly, given that the constructed dialog task in essence asked students to converse/debate, paradoxically, with/against themselves, students might not see the point of meta-talk and questioning units; in other words, there was no need to apply "epistemic vigilance" (Sperber et al., 2010) when no actual epistemic

conflict was present. By contrast, in social dialogs, the need to engage in meta-talk arises when one party wants to clarify, highlight, negotiate, or change the norms or rules of a discourse, that is to establish inter-subjectivity at the meta-level. Additionally, though thinking has been characterized as dialogic in nature as a talk with a generalized other (Graff, 2003), metacognitive awareness does not always manifest itself explicitly as verbal narratives about the process of thinking (or dialog), especially in individual reasoning contexts. Rather, meta-level self-regulation is typically implicit, expressed through strategic applications of skills and strategies to achieve specific purposes (Iordanou & Rapanta, 2021; Kuhn, 1999; Taatgen, 2013). Manifestations of implicit self-regulation include defending one's own position adeptly through drawing on relevant evidence and /or undermining a competing alternative argument by attacking its underlying reasoning (Macagno, 2016; Kuhn, 2022). Both aspects of competence were evident in the individual discourses written by experimental group students and to a lesser extent in those written by the comparison students (see Table 9 and 15 for detailed summaries). Taken together, the above reasoning convinced us that development at the metacognitive level plays significant role in conveying the benefits of dialogic engagement to individual reasoning tasks. But future research is indeed needed to further investigate this path by externalizing the implicit self-regulation processes, possibly by making students think-aloud while composing their individual discourses.

Other proposed mechanisms of transfer of skills from the dialogic to the individual dimensions include the attainment of an "oppositional voice" which prompts the individual to consider alternative possibilities and criticism when constructing an argument (Ford & Forman, 2012). Those also include the abstraction of various argument schemes which serve to organize and structure the components of an argument (Anderson & Reznitskaya, 2002), and the learning

and transfer of various proactive executive control strategies (Nussbaum & Asterhan, 2016) which might be particularly important for argumentative writing to which executive control abilities are central (Graham, 2021).

All these above assumptions implicate the applications of metacognitive competence to varying degrees. And it is in this aspect we believe the metacognitive competence, values, and dispositions dynamic proposed by Kuhn (2022) is more comprehensive and powerful in explaining why, when, and how metacognition matters in driving the development of individual and collaborative reasoning. In her model, inhibitory control is central to metacognition that explains individual's gradual attunement of strategic performance by suppressing the use of less effective strategies while promoting the use of superior ones. Regarding the development of argumentive competence, studies adopting the microgenetic method tracing the progression of skill development have found that over time with exercises, students increased their use of more advanced strategies like counter-undermine while decreased use of less effective ones like indirect-counters (Papathomas & Kuhn, 2017). In the present study, the experimental group's post-intervention dialogs also contained less unsuccessful disruptive utterances (i.e., the unconnected units) while more advanced relevant units aligned with the purposes of argumentation (e.g., direct counters and concession units). Furthermore, metacognition is stressed equally as a disposition as it is a competence. The disparities between participants' dialogic and individual writing performances observed in the present study and others (Hemberger et al., 2017; Kuhn & Moore, 2015; Shi, 2020) as well as the limited and selective transfer of skills from the social to the individual reasoning tasks (Kuhn, 2019) foreground disposition as an essential factor in determining individual performance above and beyond competence per se (Mercier et al., 2017).

4.3 Educational Implications

A Student-centered Argumentation Curriculum

In terms of its educational implications, the present study weighs in on the interest in discourse as a promising and fruitful pathway for developing individual argument skills (Rapanta & Felton, 2021; Resnick et al., 2015, 2018). Specifically, it builds upon and extends an emerging body of research focusing on apprenticeship as a developmental mechanism within the context of dialogic teaching/learning (Halpern, 2022; Mayweg-Paus et al., 2015; Papthomas & Kuhn, 2017). During one-on-one electronic argumentation sessions, anonymous experts offered individualized scaffolding to the younger participants' skills by probing the weaknesses of their arguments and reasoning (through questioning and meta-talk units) and modelling advanced counterargument strategies as needed. The results showed that the younger participants were responsive to and interiorized such implicit support by accommodating their behaviors toward the directions modelled by the experts when they were later asked to express themselves individually in writing.

Even with these positive results, it's natural to question the economical value of the dialogic approach in comparison to direct instructions because the former demands more time from the students and teachers and greater skills and resources from the teacher to manage a talking classroom. Broadly speaking, abundant research has established the advantage of dialogic teaching/learning in generating deeper understandings, broader transfer, and greater metacognitive, motivational, and even social gains in students, compared with methods characterized by direct instructions (Iordanou & Constantinou, 2015; Nussbaum & Asterhan, 2017; Resnick et al., 2015, 2018; Sun et al., 2017). The merits of the dialogic approach lie in its abilities to tap into the metacognitive venue, as discussed above, as well as to situate learning

within a broader goal-based context to stimulate and promote associative learning and transfer. Increasingly, the development of argument skill, among other higher-order cognitive competencies, has been recognized not as a single accomplishment but an integral part of some interconnected systems like critical or scientific thinking (Kuhn 2010, 2018). Argumentation, as argued by Kuhn (2018), should be integrated into and identified as the outcome component of critical thinking. It relies on and expands upon the skills and products involved in the preceding exploratory and analytical phase of inquiry learning. On this perspective, the development of argumentation should be contextualized within a broader system where its purposes and values are made clear in relation to other elements of the system, (e.g., data collection, organization, and interpretation) as well as the pragmatic goal that are shared by and unites them (Kuhn & Modreck, 2022). In this respect, direct instruction of expository writing is unduly narrow as it always targets only an isolated slate of the writing process like grammar or sentence construction while leaving others, especially those at the metacognitive and dispositional levels, unattended to (Graham & Perrin, 2007; Kuhn & Dean, 2005).

In terms of the implementations of dialogic teaching, many whole-class and group-level discussions depend on teacher's skills and resources as the discourse moderator and facilitator to be productive (Resnick et al., 2018; Reznitskaya & Wilkinson, 2017). As a result, students' outcomes can hardly be guaranteed, depending to a large extent on the skills and willingness/motivations of the teachers to take up the dialogic approach, which some of them never develop confidence or become comfortable in implementing. By contrast, the AWM program drastically reduced the responsibility of the teacher to manage the classroom discussions. The effectiveness of this method depends on the establishment of intersubjectivity and accountability among students in the process of they directly conversing with one another

and constructing norms of argumentation to which they are all expected and socially regulated to abide. During this process, teacher's input and intervention is minimized to leave room for students to observe and discover the values of new strategies and develop skills and dispositions to practice them. The activities involved in the AWM program are carefully designed and sequenced, each with clearly defined goals building on the preceding phases while preparing for the following ones so that students are clear of their purposes. In addition, by rotating participants' discourse partners at each dialogic session, we also ensure that they encounter multiple and diverse arguments from different people, a condition deemed critical to foster the motivations and competence to engage in individual argumentative reasoning (Mercier et al., 2017).

Teachers' managerial responsibility is minimized but their significance cannot be overlooked. They are critical for this kind of discovery learning to materialize. During the learning process, teachers do not directly participate in students' discussions, but they are responsible for creating a supportive environment where alternative ideas are respected, encouraged, and deliberated deeply with evidence. The creation and promotion of a communal culture by the teacher that values argumentation as a path to knowledge support students' engagement in and development of argumentation. As long as students were given adequate time to participate in it, the AWM program has consistently generated positive outcomes in its many implementations across different cultural context (see Iordanou & Rapanta, 2021 for a review). To scale up its effects, the collaboration of classroom teachers to incorporate it into regular school curriculum is key, but at least they can be assured that this would not bring about extra burden on them.

Regarding the expert component that has been found effective in the present study, additional research is needed to delineate the dimensions of expertise to promote its practical applications. But as discussed above, we believe meta-strategic scaffolding regarding the purposes and norms of arguments constitute key components of apprenticeship learning. Questioning and meta-talk moves enabled experts to implore the novices to reflect on, revise, and refine their arguments. Shi (2019) seek to generate similar meta-level support by having students fill out reflection sheets answering to questions asking them to reflect on their use of evidence during the electronic dialog sessions. She found such questioning scaffolding effective in promoting students' incorporations of belief-incongruent evidence into their arguments. Thus, certain aspects of the apprenticeship model can be realized or simulated by a more cost-effective method in the classrooms like responding to (meta-level) reflective questions. But the flexibility and capability of a real dialogic partner to support students exactly where they are struggling in the process of skill development can hardly be replaced by a less adaptable agent (Arvidsson & Kuhn, 202l).

Finally, the present study carries cultural significance in the sense that it demonstrates that the apprenticeship approach is also effective in culturally eastern contexts (Henrich et al., 2010) where the values of affiliation, harmony, and avoidance of conflict, rather than argumentation, are emphasized and rehearsed in many spheres of the society. Students from culturally western contexts in general have had a head start in developing argument skills because they are exposed to and (thus) familiar with argumentation as a cultural and historical tool to advance knowledge and solve (controversial) social/political issues since very young. By contrast, students from culturally eastern contexts like China are traditionally expected and disciplined to obey authoritative figures at home and in schools and withhold their personal opinions that conflict

with those of the authorities. Argumentation, especially those occurring in public contexts, are despised as reflections of bad disciplines of its participants and morally discouraged when it involves arguing with an elder (e.g., one's parent or teacher) in China. Thus, it's reasonable to question the affordances of the argumentation-based curriculum when it is to be used in a context where its main activity seems to be at odds with the culturally applauded behaviors. Our observation that at pretest students from both conditions showed negligible dialogic argumentation skill corresponds to the natural inference one would make based on this stereotypical description. However, the fact that over time these Chinese adolescents grew in competencies to argue with an opposing side to support their own positions shows that they were capable of adapting to and profiting from the discourse-based argumentative curriculum. They showed similar trajectories of skill development as those demonstrated by US adolescents who underwent similar argumentation-based curriculum (2016Halpern, 2022; Papathomas & Kuhn,), despite any culturally rooted inclination of them to avoid conflict and maintain harmonious relationships (Kuhn et al., 2011). Our demonstration that the effects were more profound in the experimental condition further defies the obsolete assumption that Chinese students can only master ready-made knowledge prepared by their teachers. On the contrary, they were able to recognize more advanced argumentive strategies modelled by "experts" and modify their behaviors in the direction of more sophisticated arguments in a self-directed manner.

In a previous implementation of the argumentation curriculum with older Chinese adolescents (i.e., 7th and 9th graders) from an economically more affluent background, Shi (2019, 2020a) found that her students demonstrated elevated levels of usages of functional and evidence-based types of arguments by the time of posttest, relative to their pretest performance. Her students showed more counterargument and evidence-based units in their argumentation,

compared to the levels of usage exhibited by students in the present study by the time of posttests. The facts that the two groups of students started their programs with diverse initial skill levels (due to age and other socioeconomic factors) and underwent different types and lengths of interventions (one and a half months in the present study versus four months in Shi's studies) made a direct comparison of the two studies impossible. Instead of applying the apprenticeship approach, Shi attempted to augment students' performance by prompting them to reflect on their use of various types of evidence by having them fill out evidence reflection sheets predicting and evaluating their use of evidence during electronic dialog sessions. This added reflective activity was proven effective and the positive results were attributed to development at the meta-strategic level regarding the functions and values of evidence in arguments. Thus, her and the present studies did converge on the significance of meta-level scaffolding that prompts students to reflect on and improve their arguments on the basis of an enhanced understanding of the functions and values of (various components of) an argument.

4.4 Limitations and future directions

Our results in general confirmed the hypothesis we set out to test, that is the effect of the apprenticeship model for developing individual argumentive skills within a dialogic argumentation context. However, the quest does not stop here, rather, more questions arise worth investigating to advance our understandings of the apprenticeship mechanism and the dialogic approach in order to promote their practical applications.

First of all, the expertise dimension needs more delineation in terms of its parameters and operationalizations in practice. Our results showed that emerging adolescents were capable of recognizing more advanced skills and accommodating their behaviors according to higher

standards under anonymous situations, that is in the absence of explicit social cues. This echoes with Fedyke and colleagues' (2017) assumption of children's developing epistemological theory of evidence that enabled them to identify expertise in relation to the goals of tasks. However, in reality, more often than not children's interactions with more skilled others (e.g., parents, teachers, and older peers) are off-line and direct, or at least with each other's social statuses/identities known. This raises the question of whether, and to what extent, can our results generalize into more conventional social contexts where various social cues are at play and might affect children's judgments of expertise and competence? Would conversing face-to-face with an adult lead to premature closure of a child's thinking because of the pressure of adult authority, as suggested by Piaget, or result in more attention to the communicated knowledge from and observed behavior of the adult due to trust and respect? What's the optimal age or skill differences between a novice and an expert taking into account the effects of social status disparities and qualities of collaboration? Would the demographic and socioeconomic status of an adult in relation to that of a child impact the child' attitudes toward the expert and subsequent learning outcomes? Would gains have been greater if a child exclusively interacted with adults? Would the modes of communication matter, for example, would face-to-face/instant dialog generate greater gains than text-mediated/delayed communications? And would the perceived (learning) history (e.g., the learning outcomes and processes) of an adult impacts a child's perception of their expertise and dispositions to learn from them? These questions and many others await further investigation so that we could design the optimal collaborative learning conditions.

Another limitation of the present study concerns with the order of the two posttest assessments. The dialog construction task was assigned first, followed by the writing task.

However, there has been disagreement regarding whether the order of presentations (of dialog and writing tasks) would impact students' essay performance. Zavala and Kuhn (2017) found that the constructed dialog task prompted college students to represent and compare two contrasting positions more adequately with reference to evidence in comparison with the non-dialogic essay writing task. This difference subsequently led to significant disparities in the qualities of students' individual argumentive writing that defends their favored position. By contrast, Macagno (2016) compared the qualities of evidence use of two groups of high school students who engaged in both a dialog task and an essay writing task but in reverse orders and found that the order of tasks had no impact on students' performance. In the present study, we used different coding schemes to analyze dialogs and essays, and this enabled us to capture the unique and essential features of both modes of communication (Kuhn, 2019; Shi, 2020a) and to infer about students' understandings of them as discussed in depth above. However, this also largely constrained our ability to directly compare students' dialog and essay performance as most codes did not overlap. Thus, we do not make any assumption here regarding the order effect. Regarding the internal validity of our results, we do not consider that the order of presentation posed an existential threat to our interpretations. The rationales are as follows: if engaging in an (imagined) dialog does prompt the students to represent and process opposing positions better which helps with their essay writing as suggested by Zavala and Kuhn (2017), our results that experimental group outperformed the comparison group in essays in spite of this shared priming effects (i.e., an equalizer) should be an underestimate of the real power of the apprenticeship model, that is the observed effects of discoursing with experts during the intervention should have been larger if students didn't take the "equalizing" dialog task first. If the alleged priming effect of dialog does not exist as observed by Macagno (2016), the order of

the presentations of the two tasks has little impact on our assessment of the efficacy of the apprenticeship mechanism. Regardless, whether or not engaging in a dialog (either socially or individually) mediates the effects of argumentation on individual essay writing remains an open question. Future research could further investigate this question by varying the order of presentations and developing coding schemes that could capture shared skills involved in both modalities of communications with diverse populations of different prior knowledge and skill levels.

Thirdly, it should be noted that our assumption that students who had the opportunity of arguing with experts developed better metacognitive competence and dispositions that supported their subsequent individual reasoning tasks was only speculative, even though our findings at the performance level did suggest their enhanced metacognitive awareness of and regulation in accordance with the purposes of arguments and argumentation modelled by the experts. Thus, an important next step is to uncover students' metacognitive regulation when they reason together and alone. Are students metacognitively aware of and motivated to interiorized advanced strategies? Are they epistemologically active and updating their understandings of the purposes of arguments and evidence in the process of dialogs? Do they engage in metacognitive planning and monitoring when writing and if so, how do they achieve this?

Relatedly, another important factor that impacts students' performance on reasoning tasks, besides competence, is their motivations (Kuhn, 2019). To the extent that dispositions to exercise a certain strategy/skill play a role in performance, what students demonstrate in their school assignments should not be taken for granted as their typical practices in their daily lives outside of schools. The school context is argued to elicit a school-related genre of writing which is characterized by students' tendency to write exclusively for the teacher by avoiding any

mention of arguments that might suggest against their central claim (Kuhn & Moore, 2015). While outside of school, students might feel more freedom and be more open to various opinions. The so-called school-related genre, motivated by students' inclination to write to the teacher, could be invoked to partially explain why students still showed my-side bias in their essays even after participating in the AWM program for extended periods as evidenced in the present as well as many other studies (; Hemberger et al., 2017; Kuhn & Moore, 2015Shi, 2019). However, the competitive environment of school might also motivated students to perform as best as they are capable of while their everyday reasoning practices tended to be uncritical and lazy (Mercier et al., 2017). In a word, we still lack a clear understanding regarding the transfer of reasoning skills between educational and more informal settings. Future research could contribute to this question by studying argumentive reasoning in diverse contexts and understand what factors contribute to the dispositions to exercise it, as opposed to competence.

Finally, with the expanding influence of the (rational) constructivism, educational and psychological researchers and practitioners are paying increasing attention to students' motivational and epistemological profiles to explain why and how reasoning development does or does not happen. This is, without question, a productive path to pursue. However, as a sociocultural endeavor, the development of argumentative reasoning inevitably entails a bi-directional process between the one who learns and the one who teaches (Zillmer & Kuhn, 2018). This means the dispositions of the teachers also matter, besides those of the students. Most fundamentally, the development of argumentive skills, as demonstrated by the present study, should be embedded within a supportive and nurturing environment that values and provides opportunities for students to exercise them. The transition from a traditional teacher-centered classroom to a student-centered dialogic one exacts a considerable challenge for the

teachers. This transition predicates on the shifting of authority of knowledge from the teacher to evidence and argument (Rapanta, 2019). The latter is no longer a priori socially imposed authority as is the case for the former, rather it is founded on the shared construction of meaning (Kuhn, 2019). The present as well as many other studies have suggested that this shift of authority on the part of the students can only be accomplished through sustained engagement in discourse with a clear epistemic purpose (Kuhn & Zillmer, 2015; Kuhn et al., 2013; Shi, 2020b). Though research has demonstrated that teachers' epistemic beliefs impact their willingness to adopt the dialogic approach (Klieme, et al., 2009; Pauli & Reusser, 2015;), knowledge regarding how to update and develop teacher's motivational and epistemic profiles to support their adoption and implementation of it is lacking. The present dialogic intervention was implemented mainly by our research team with very limited involvement of the classroom teachers to keep students disciplined in the process. Anecdotal evidence from interacting with these teachers suggested to us that though they were curious about this novel approach, they were not fully convinced about its advantages over the more traditional authoritative method especially given the pressure to produce the best test results within limited timeframes. Thus, to scale up instructional revolution toward discourse-based student-centered classrooms, the motivations of front-line teachers must be prioritized. They must be convinced about the values of this change, committed to the merits of discourse both as a means and an educational objective in itself (Kuhn, 2019). This is not an easy task especially given the misalignments among state-level standards, the contents of widely endorsed assessment tools, and the entrenched pedagogical beliefs dominating our current educational system (Resnick et al., 2018). Continued efforts should be made to enhance the agencies and responsibilities of both students and teachers to

contribute to the shared endeavor of knowing and learning as sociocultural and historical practices.

References

Anderson, R. C., Nguyen-Jahiel, K., McNurlen, B., Archodidou, A., Kim, S., Reznitskaya, A., Tillmanns, M., & Gilbert, L. (2001). The snowball phenomenon: Spread of ways of talking and ways of thinking across groups of children. *Cognition and Instruction*, *19*(1), 1–46. https://doi.org/10.1207/S1532690XCI1901_1

Andriessen, J.E.B., Baker, M., Suthers, D., Andriessen, J., Baker, M., Suthers, D., Universiteit Utrecht, & Afd Educatie. (2003). Argumentation, computer support, and the educational context of confronting cognitions. In *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments* (pp. 1–25). Kluwer. https://dspace.library.uu.nl/handle/1874/27661

Arvidsson, T. S., & Kuhn, D. (2021). Realizing the full potential of individualizing learning. *Contemporary Educational Psychology*, *65*, 101960. https://doi.org/10.1016/j.cedpsych.2021.101960

Asterhan, C. S. C., & Schwarz, B. 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

Baines, E., Rubie-Davies, C., & Blatchford, P. (2009). Improving pupil group work interaction and dialogue in primary classrooms: Results from a year-long intervention study. *Cambridge Journal of Education*, *39*(1), 95–117. https://doi.org/10.1080/03057640802701960

Billig, M. (1987). *Arguing and thinking: A rhetorical approach to social psychology*. Cambridge University Press.

Bridgers, S., De Simone, C., Gweon, H., & Ruggeri, A. (2023). Children seek help based on how others learn. *Child Development*, *00*, 1–22. https://doi.org/10.1111/cdev.13926

Common Core State Standards Initiative. (2010). Common core state standards for English language arts and literacy in history/social studies, science, and technical subjects. Retrieved from http://www.corestandards.org/assets/CCSSI_ELA%20Standards.pdf

Cross, D., Taasoobshirazi, G., Hendricks, S., & Hickey, D. T. (2008). Argumentation: A strategy for improving achievement and revealing scientific identities. *International Journal of Science Education*, *30*(6), 837–861. https://doi.org/10.1080/09500690701411567

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

Damico, J. S., & Rosaen, C. L. (2009). Creating epistemological pathways to a critical citizenry: Examination of a fifth-grade discussion of freedom. *Teachers College Record*, *111*(5), 1163–1194. https://doi.org/10.1177/016146810911100501

Fedyk , M. & Xu , F. (2017). Th epistemology of rational constructivism . *Review of Philosophy and Psychology , 9,* 343-362.

Fedyk, M., Kushnir, T., & Xu, F. (2019). Intuitive epistemology: Children's theory of evidence. In D. Wilkenfeld & R. Samuels (Eds.), *Advances in Experimental Philosophy of Science*. London: Bloomsbery.

Felton, M., & Kuhn, D. (2001). The development of argumentive discourse skill. *Discourse Processes*, *32*(2–3), 135–153. https://doi.org/10.1080/0163853X.2001.9651595

Felton, M., Crowell, A., & Liu, T. (2015). Arguing to agree: Mitigating my-Side bias through consensus-seeking dialogue. *Written Communication*, *32*(3), 317–331. https://doi.org/10.1177/0741088315590788

Ford, M. J. (2012). A dialogic account of sense-making in scientific argumentation and reasoning. *Cognition and Instruction*, *30*(3), 207–245.

Ford, M. J., & Forman, E. A. (2015). Uncertainty and scientific progress in classroom dialogue. In L. B. Resnick, C. S. C. Asterhan, & S. N. Clarke (Eds.), *Socializing Intelligence Through Academic Talk and Dialogue* (pp. 143–155). https://doi.org/10.3102/978-0-935302-43-1_11

Forman, E. A. (1987). Learning through peer interaction: A Vygotskian perspective. *The Genetic Epistemologist*, *15,* 6-15.

Gergen, K. J. (1992). The social constructionist movement in modern psychology. In *The restoration of dialogue: Readings in the philosophy of clinical psychology* (pp. 556–569). American Psychological Association. https://doi.org/10.1037/10112-044

Golanics, J. d., & Nussbaum, E. m. (2008). Enhancing online collaborative argumentation through question elaboration and goal instructions. *Journal of Computer Assisted Learning*, *24*(3), 167–180. https://doi.org/10.1111/j.1365-2729.2007.00251.x

Graff, G. (2003). *Clueless in academe: How schooling obscures the life of the mind*. Yale University Press. https://www.jstor.org/stable/j.ctt1npkd5

Graham, S. (2021). Executive control and the writer(s)-within-community model. In S. Graham, *Executive Functions and Writing* (pp. 38–76). Oxford University Press. https://doi.org/10.1093/oso/9780198863564.003.0003

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

Graham, S., & Rijlaarsdam, G. (2016). Writing education around the globe: Introduction and call for a new global analysis. *Reading and Writing*, *29*(5), 781–792. https://doi.org/10.1007/s11145-016-9640-1

Grice, H. P. (1975). Logic and conversation. In *Speech Acts* (pp. 41–58). Brill. https://doi.org/10.1163/9789004368811_003

Ha, W., & Yu, R. (2019). Quasi-experimental evidence of a school equalization reform on housing prices in beijing. *Chinese Education & Society*, *52*(3–4), 162–185. https://doi.org/10.1080/10611932.2019.1667681

Halpern, M. R. (2022). *Youth apprenticeship in reasoned discourse: The power of learning by doing* [Ph.D., Columbia University]. https://www.proquest.com/docview/2607569457/abstract/BD11F3B47F424CB5PQ/1

Harris, P. (2012). *Trusting what you're told: How children learn from others*. Cambridge MA: Harvard University Press.

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

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge University Press.

Howe, C., & Zachariou, A. (2019). Small-group collaboration and individual knowledge acquisition: The processes of growth during adolescence and early adulthood. *Learning and Instruction*, *60*, 263–274. https://doi.org/10.1016/j.learninstruc.2017.10.007

Hsu, P.-S., Van Dyke, M., Chen, Y., & Smith, T. j. (2015). The effect of a graph-oriented computer-assisted project-based learning environment on argumentation skills. *Journal of Computer Assisted Learning*, *31*(1), 32–58. https://doi.org/10.1111/jcal.12080

Iordanou, K. A. (2010). *Developing argument skills across scientific and social domains* [Ph.D., Columbia University]. https://www.proquest.com/docview/304652301/abstract/4FCE347ABF4F13PQ/1

Iordanou, K. (2022). Supporting strategic and meta-strategic development of argument skill: The role of reflection. *Metacognition and Learning*, *17*(2), 399–425. https://doi.org/10.1007/s11409-021-09289-1

Iordanou, K., & Constantinou, C. P. (2015). Supporting use of evidence in argumentation through practice in argumentation and reflection in the context of SOCRATES learning environment. *Science Education*, *99*(2), 282–311. https://doi.org/10.1002/sce.21152

Iordanou, K., Kuhn, D., Matos, F., Shi, Y., & Hemberger, L. (2019). Learning by arguing. *Learning and Instruction*, *63*, 101207. https://doi.org/10.1016/j.learninstruc.2019.05.004

Iordanou, K., & Rapanta, C. (2021). "Argue with me": A method for developing argument skills. *Frontiers in Psychology*, *12*. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.631203

Jin, Z., Wang, X., & Huang, B. (2023). The enrolment reform of schools and housing price: Empirical evidence from Shanghai, China. *International Review of Economics & Finance*, *84*, 262–273. https://doi.org/10.1016/j.iref.2022.11.024

Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms. *The power of video studies in investigating teaching and learning in the classroom*, *137* - 160.

Koschmann, T. (1999). *Toward a dialogic theory of learning: Bakhtin's contribution to understanding learning in settings of collaboration*.

Krabbe, E. C. W. (2003). Metadialogues. In F. H. Van Eemeren, J. A. Blair, C. A. Willard, & A. F. Snoeck Henkemans (Eds.), *Anyone who has a view: Theoretical contributions to the study of argumentation* (pp. 83–90). Springer Netherlands. https://doi.org/10.1007/978-94-007-1078-8_7

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

Kuhn, D. (1995). Microgenetic study of change: What has it told us? *Psychological Science*, *6*(3), 133–139. https://doi.org/10.1111/j.1467-9280.1995.tb00322.x

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

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

Kuhn, D. (2010). Teaching and learning science as argument. *Science Education*, *94*(5), 810–824. https://doi.org/10.1002/sce.20395

Kuhn, D. (2015). Thinking together and alone. *Educational Researcher*, *44*(1), 46–53. https://doi.org/10.3102/0013189X15569530

Kuhn, D. (2018). A role for reasoning in a dialogic approach to critical thinking. *Topoi*, *37*(1), 121–128. https://doi.org/10.1007/s11245-016-9373-4

Kuhn, D. (2019). Critical thinking as discourse. *Human Development*, *62*(3), 146–164. https://doi.org/10.1159/000500171

Kuhn, D. (2022). Metacognition matters in many ways. *Educational Psychologist*, *57*, 1–14. https://doi.org/10.1080/00461520.2021.1988603

Kuhn, D., & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents' thinking. *Psychological Science*, *22*(4), 545–552. https://doi.org/10.1177/0956797611402512

Kuhn, D., & Dean, D. (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science*, *16*(11), 866–870. https://doi.org/10.1111/j.1467-9280.2005.01628.x

Kuhn, D., & Modrek, A. S. (2022). Choose your evidence: Scientific thinking where it may most count. *Science & Education*, *31*(1), 21–31. https://doi.org/10.1007/s11191-021-00209-y

Kuhn, D., & Moore, W. (2015). Argumentation as core curriculum. *Learning: Research and Practice*, *1*(1), 66–78. https://doi.org/10.1080/23735082.2015.994254

Kuhn, D., & Zillmer, N. (2015). Developing norms of discourse. In L. B. Resnick, C. S. C. Asterhan, & S. N. Clarke (Eds.), *Socializing intelligence through academic talk and dialogue* (pp. 77–86). https://doi.org/10.3102/978-0-935302-43-1_6

Kuhn, D., Halpern, M., & Bruun, S. (2022). From talk to text: Implementing student discussions that matter. In Z. Philippakos & S. Graham (Eds.), *Writing and reading connections: Bridging research and practice. Guilford.*

Kuhn, D., Hemberger, L., & Khait, V. (2016). Tracing the development of argumentive writing in a discourse-rich context. *Written Communication*, *33*(1), 92–121. https://doi.org/10.1177/0741088315617157

Kuhn, D., Wang, Y., & Li, H. (2011). Why argue? Developing understanding of the purposes and values of argumentive discourse. *Discourse Processes*, *48*(1), 26–49. https://doi.org/10.1080/01638531003653344

Kuhn, D., Zillmer, N., Crowell, A., & Zavala, J. (2013). Developing norms of argumentation: Metacognitive, epistemological, and social dimensions of developing argumentive competence. *Cognition and Instruction*, *31*(4), 456–496. https://doi.org/10.1080/07370008.2013.830618

Macagno, F. (2016). Argument relevance and structure. Assessing and developing students' uses of evidence. *International Journal of Educational Research*, *79*, 180–194. https://doi.org/10.1016/j.ijer.2016.07.002

Matos, F. (2021). Collaborative writing as a bridge from peer discourse to individual argumentative writing. *Reading and Writing*, *34*(5), 1321–1342. https://doi.org/10.1007/s11145-020-10117-2

Mayweg-Paus, E., Macagno, F., & Kuhn, D. (2015). Developing argumentation strategies in electronic dialogs: Is modeling effective? *Discourse Processes*, *53*(4), 280–297. https://doi.org/10.1080/0163853X.2015.1040323

Mead, G. H. (1934). *Mind, self, and society from the standpoint of a social behaviorist*. Chicago.

Mercer, N. (1995). *The guided construction of knowledge: Talk amongst teachers and learners*. Multilingual Matters.

Mercer, N. (2004). Sociocultural discourse analysis. *Journal of Applied Linguistics, 1*(2), 137–168. https://doi. org/10.1558/japl.v1i2.137

Mercer, N., Hennessy, S., & Warwick, P. (2017). Dialogue, thinking together and digital technology in the classroom: Some educational implications of a continuing line of inquiry. *International Journal of Educational Research*, *97*, 187–199. https://doi.org/10.1016/j.ijer.2017.08.007

Mercer, N., & Howe, C. (2012). Explaining the dialogic processes of teaching and learning: The value and potential of sociocultural theory. *Learning, Culture and Social Interaction*, *1*(1), 12–21. https://doi.org/10.1016/j.lcsi.2012.03.001

Mercier, H., Boudry, M., Paglieri, F., & Trouche, E. (2016). Natural-born arguers: Teaching how to make the best of our reasoning abilities. *Educational Psychologist*, *52*, 1–16. https://doi.org/10.1080/00461520.2016.1207537

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(2), 57–74. https://doi.org/10.1017/S0140525X10000968

Muller Mirza, N., Tartas, V., Perret-Clermont, A.-N., & de Pietro, J.-F. (2007). Using graphical tools in a phased activity for enhancing dialogical skills: An example with Digalo. *International Journal of Computer-Supported Collaborative Learning*, *2*(2), 247–272. https://doi.org/10.1007/s11412-007-9021-5

Next Generation Science Standards: For States, By States. (2013). National Academies Press. https://doi.org/10.17226/18290

Nussbaum, E. M. (2008). Collaborative discourse, argumentation, and learning: Preface and literature review. *Contemporary Educational Psychology*, *33*(3), 345–359. https://doi.org/10.1016/j.cedpsych.2008.06.001

Nussbaum, E. M. (2021). Critical integrative argumentation: Toward complexity in students' thinking. *Educational Psychologist*, *56*(1), 1–17. https://doi.org/10.1080/00461520.2020.1845173

Nussbaum, E. M., & Asterhan, C. S. C. (2016). *The psychology of far transfer from classroom argumentation*.

OECD. (2020). *Benchmarking the performance of China's education system*. OECD. https://doi.org/10.1787/4ab33702-en

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

Pauli, C., & Reusser, K. (2015). Discursive cultures of learning in (everyday) mathematics teaching: A video-based study on mathematics teaching in German and Swiss classrooms. *Socializing intelligence through academic talk and dialogue*, 181-193.

Piaget, J. (1976). Piaget's Theory. In B. Inhelder, H. H. Chipman, & C. Zwingmann (Eds.), *Piaget and his School: A reader in developmental psychology* (pp. 11–23). Springer. https://doi.org/10.1007/978-3-642-46323-5_2

Piaget, J., & lnhelder, B. *The psychology of the child.* New York: Basic Books, 1969. (Original work published in 1966)

Rapanta, C. (2019). Argumentation as critically oriented pedagogical dialogue. *Informal Logic*, *39*(1), 1–31. https://doi.org/10.22329/il.v39i1.5116

Rapanta, C., & Felton, M. K. (2021). Learning to argue through dialogue: A review of instructional approaches. *Educational Psychology Review*, *34*(2), 477–509. https://doi.org/10.1007/s10648-021-09637-2

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

Resnick, L. B., Asterhan, C. S., Clarke, S. N., & Schantz, F. (2018). Next generation research in dialogic learning. *Wiley Handbook of Teaching and Learning*, 323–338.

Reznitskaya, A., & Gregory, M. (2013). Student thought and classroom language: Examining the mechanisms of change in dialogic teaching. *Educational Psychologist*, *48*(2), 114–133. https://doi.org/10.1080/00461520.2013.775898

Reznitskaya, A., & Wilkinson, I. A. G. (2017). Truth matters: Teaching young students to search for the most reasonable answer. *Phi Delta Kappan*, *99*(4), 33–38. https://doi.org/10.1177/0031721717745550

Reznitskaya, A., Anderson, R. C., & Kuo, L. (2007). Teaching and learning argumentation. *The Elementary School Journal*, *107*(5), 449–472. https://doi.org/10.1086/518623

Reznitskaya, A., Glina, M., Carolan, B., Michaud, O., Rogers, J., & Sequeira, L. (2012). Examining transfer effects from dialogic discussions to new tasks and contexts. *Contemporary Educational Psychology*, *37*(4), 288–306. https://doi.org/10.1016/j.cedpsych.2012.02.003

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context* (pp. xiv, 242). Oxford University Press.

Rogoff, B. (1995). Observing sociocultural activity on three planes: Participatory appropriation, guided participation, and apprenticeship. In *Sociocultural studies of mind* (pp. 139–164). Cambridge University Press. https://doi.org/10.1017/CBO9781139174299.008

Ryu, S., & Sandoval, W. A. (2012). Improvements to elementary children's epistemic understanding from sustained argumentation. *Science Education*, *96*(3), 488–526. https://doi.org/10.1002/sce.21006

Sampson, V., & Clark, D. B. (2011). A comparison of the collaborative scientific argumentation practices of two high and two low Performing groups. *Research in Science Education*, *41*(1), 63–97. https://doi.org/10.1007/s11165-009-9146-9

Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction*, *23*(1), 23–55. https://doi.org/10.1207/s1532690xci2301_2

Scribner, S., & Cole, M. (2013). The psychology of literacy. In *The Psychology of Literacy*. Harvard University Press. https://doi.org/10.4159/harvard.9780674433014

Sfard, A. (2015). Why all this talk about talking classrooms? Theorizing the relation between talking and learning. In L. B. Resnick, C. S. C. Asterhan, & S. N. Clarke (Eds.), *Socializing Intelligence Through Academic Talk and Dialogue* (pp. 245–253). https://doi.org/10.3102/978-0-935302-43-1_19

Shi, Y. (2019). Enhancing evidence-based argumentation in a Mainland China middle school. *Contemporary Educational Psychology*, *59*, 101809. https://doi.org/10.1016/j.cedpsych.2019.101809

Shi, Y. (2020a). Constructed dialogs reveal skill development in argumentive writing. *Reading and Writing*, *33*(9), 2311–2335. https://doi.org/10.1007/s11145-020-10045-1

Shi, Y. (2020b). Talk about evidence during argumentation. *Discourse Processes*, *57*(9), 770–792. https://doi.org/10.1080/0163853X.2020.1777498

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

Shvarts, A., & Bakker, A. (2019). The early history of the scaffolding metaphor: Bernstein, Luria, Vygotsky, and before. *Mind, Culture, and Activity*, *26*(1), 4–23. https://doi.org/10.1080/10749039.2019.1574306

Siegler, R. S. (2006). Microgenetic analyses of learning. In *Handbook of child psychology: Cognition, perception, and language, Vol. 2, 6th ed* (pp. 464–510). John Wiley & Sons, Inc.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, *25*(4), 359–393. https://doi.org/10.1111/j.1468-0017.2010.01394.x

Sun, J., Anderson, R. C., Perry, M., & Lin, T.-J. (2017). Emergent leadership in children's cooperative problem solving groups. *Cognition and Instruction*, *35*(3), 212–235. https://doi.org/10.1080/07370008.2017.1313615

Taatgen, N. A. (2013). The nature and transfer of cognitive skills. *Psychological Review*, *120*(3), 439–471. https://doi.org/10.1037/a0033138

Toulmin, S. E. (2003). *The uses of argument* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511840005

Tudge, J. R. H., & Winterhoff, P. A. (1993). Vygotsky, Piaget, and Bandura: Perspectives on the relations between the social world and cognitive development. *Human Development*, *36*(2), 61–81. https://doi.org/10.1159/000277297

Van Eemeren, F. H., & Grootendorst, R. (1992). Relevance reviewed: The case of argumentum ad hominem. *Argumentation*, *6*(2), 141–159. https://doi.org/10.1007/BF00154322

Vygotsky, L. S. (1987). *The collected works of L.S. Vygotsky: The fundamentals of defectology*. Springer Science & Business Media.

Vygotsky, L. S., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press.

Walton, D. (1989). *Informal logic: A handbook for critical argument*. Cambridge University Press.

Walton, D. (2014). *Abductive reasoning*. University of Alabama Press.

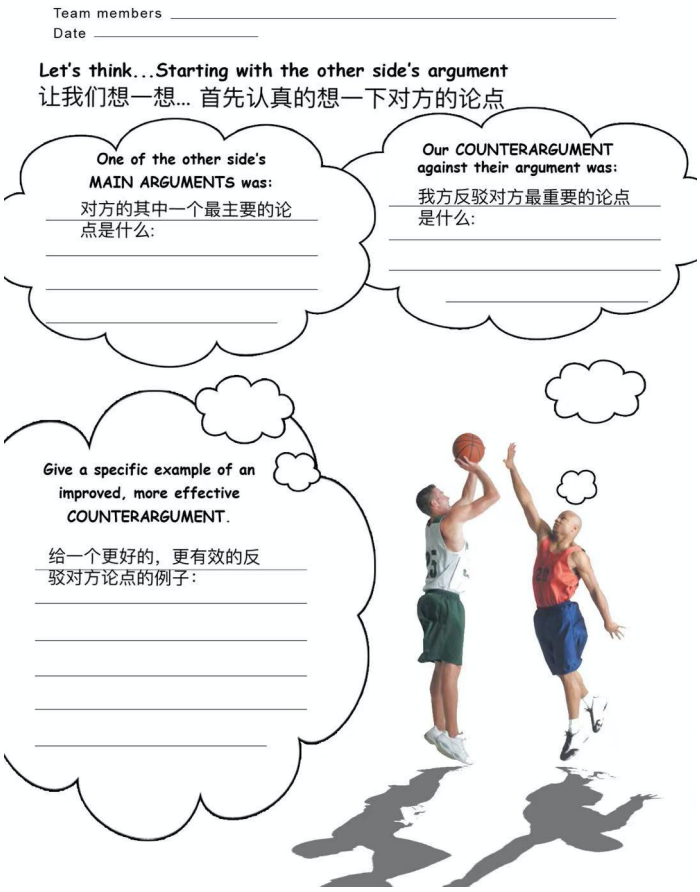
Zavala, J., & Kuhn, D. (2017). Solitary discourse is a productive activity. *Psychological Science*, *28*(5), 578–586. https://doi.org/10.1177/0956797616689248

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

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*(2), 172–223. https://doi.org/10.1016/j.dr.2006.12.001

Appendix A

Reflection Forms on the Opponent- and Own-side arguments



Team members _____

Date _____

Let's think...Starting with our argument 让我们想一想..... 首先认真的想一想我方论点

One of our MAIN ARGUMENTS was:
我方最重要的一个论点是什 么:

Their COUNTERARGUMENT against our argument was:
对方反驳我方最重要的论点是什么:

Our COMEBACK was:
我方做出的回应是什么:

How can this COMEBACK be improved? Is there a more effective comeback?
这个回应怎么样可以变的更好? 你们可以想到一个更有效的回应吗?

Appendix B

Questions & Answers for Soda Topic

Function	Position: Teens should be allowed to drink soda	Position: Teens should be restricted in soda consumption
Support my own side (M+)	**喝汽水在短期内如何影响大脑？(How does soda affect the brain in the short term?)** 当一个人摄入汽水中的糖分时，大脑中的奖励系统就会被激活。大脑会释放一种叫做多巴胺的化学物质，它会让人产生愉悦的感觉。 (When a person consumes sugar, as found in soda, the reward system in the brain is activated. The brain releases chemical called dopamine, which produces feelings of pleasure.)	**不喝汽水可以改善健康吗？(Can avoiding soda improve health?)** 不喝汽水可以通过降低患高血压、哮喘、胰腺癌和糖尿病的几率来改善健康。 (Avoiding soda can improve health by lowering the chances of hypertension, asthma, pancreatic cancer, and diabetes.)
	喝汽水对健康有好处吗？(Does drinking soda have any health benefits?) 一项针对有便秘症状的人的研究发现，喝碳酸水的人比喝普通自来水的人的便秘症状更少，而汽水中含有碳酸水. (A study of people with constipation symptoms found that those who drank carbonated water, as found in soda, had fewer symptoms than those who drank normal tap water.)	**避免喝汽水可以帮助人们减肥吗？(Can avoiding soda help people to lose weight?)** 一项研究表明，减少汽水摄入量的高中生体重保持不变。而没有改变的学生体重增加了。(A study showed that high schoolers who reduced their soda intake maintained the same weight. Students who made no change gained weight.)
Weaken opposing side (O-)	**所有类型的汽水都含有很多糖和卡路里吗？(Are all types of sodas high in sugar and calories?)**	**喝普通汽水会导致体重增加吗？(Can drinking regular soda lead to weight gain?)** 研究表明，喝汽水可能会让人想吃得更多，因为它会导

	现在有许多无糖或0卡路里的汽水。(Most sodas are available in no-calorie types.)	致血糖升高。 在一项研究中，在日常饮食中添加汽水的人多摄入了 17% 的卡路里。 摄入更多卡路里会导致体重增加。(Research suggests that drinking soda may make people want to eat more because it causes a rise in blood sugar. In one study, people who added soda to their regular diet consumed 17% more calories. Consuming more calories leads to weight gain.)
	关于汽水和肥胖的研究都是谁做的？(Who conducts most of the research on soda and obesity?) 大多数发现汽水和肥胖之间存在联系的研究都是由反对饮用汽水的研究人员进行的。(Most studies that have found links between soda and obesity have been conducted by researchers who are against consuming soda.)	**汽水中的咖啡因所提供的短期能量提升能持续多久？(How long-lasting is the short-term energy boost from the caffeine found in soda?)** 研究发现，摄入咖啡因的学生更容易晚上睡不着觉，从而早上起来更容易感到疲倦。(Research has found that students who consume caffeine are more likely to have trouble sleeping and to feel tired in the morning.)
Support opposing side (O+)	**不喝汽水可以改善健康吗？(Can avoiding soda improve health?)** 不喝汽水可以通过降低患高血压、哮喘、胰腺癌和糖尿病的几率来改善健康。(Avoiding soda can improve health by lowering the chances of hypertension, asthma, pancreatic cancer, and diabetes.)	**喝汽水在短期内如何影响大脑？(How does soda affect the brain in the short term?)** 当一个人摄入汽水中的糖分时，大脑中的奖励系统就会被激活。大脑会释放一种叫做多巴胺的化学物质，它会让人产生愉悦的感觉。(When a person consumes sugar, as found in soda, the reward system in the brain is activated. The brain releases chemical called dopamine,

		which produces feelings of pleasure.)
	避免喝汽水可以帮助人们减肥吗？(Can avoiding soda help people to lose weight?) 一项研究表明，减少汽水摄入量的高中生体重保持不变。而没有改变的学生体重增加了。(A study showed that high schoolers who reduced their soda intake maintained the same weight. Students who made no change gained weight.)	**喝汽水对健康有好处吗？(Does drinking soda have any health benefits?)** 一项针对有便秘症状的人的研究发现，喝碳酸水的人比喝普通自来水的人的便秘症状更少，而汽水中含有碳酸水. (A study of people with constipation symptoms found that those who drank carbonated water, as found in soda, had fewer symptoms than those who drank normal tap water.)
Weaken own side (M-)	**汽水中的咖啡因所提供的短期能量提升能持续多久？(How long-lasting is the short-term energy boost from the caffeine found in soda?)** 研究发现，摄入咖啡因的学生更容易晚上睡不着觉，从而早上起来更容易感到疲倦。(Research has found that students who consume caffeine are more likely to have trouble sleeping and to feel tired in the morning.)	**所有类型的汽水都含有很多糖和卡路里吗？(Are all types of sodas high in sugar and calories?)** 现在有许多无糖或0卡路里的汽水。(Most sodas are available in no-calorie types.)
	喝普通汽水会导致体重增加吗？(Can drinking regular soda lead to weight gain?) 研究表明，喝汽水可能会让人想吃得更多，因为它会导致血糖升高。 在一项研究中，在日常饮食中添加汽水的人多摄入了 17% 的卡路里。 摄入更多卡路里会导致体重增加。(Research suggests that drinking soda may make people want to eat	**关于汽水和肥胖的研究都是谁做的？(Who conducts most of the research on soda and obesity?)** 大多数发现汽水和肥胖之间存在联系的研究都是由反对饮用汽水的研究人员进行的。(Most studies that have found links between soda and obesity have been conducted by researchers who are against consuming soda.)

	more because it causes a rise in blood sugar. In one study, people who added soda to their regular diet consumed 17% more calories. Consuming more calories leads to weight gain.)	

Appendix C

Questions & Answers for Animal Research Topic

Functions	Position: Animals can be used	Position: animals cannot be used
Support my own side (M+)	问：为什么会用动物来做医学研究？**(Why have animals been used in research?)** 答：动物器官和人类器官常常非常相似，所以医药的疗效也通常很类似。(Animal organs often resemble human organs, so medicines may work in similar ways.)	问：可以用人体细胞模型来测试新药吗？（**Can information be gained from studying human cells in a laboratory?**） 答：最新开发的由人体细胞制成的模型旨在更接近地模拟人体的工作方式，并可用于测试人类对药物的反应。（Studies of human cells under a microscope provide valuable information and can avoid the need for surgery. For example, examining human tissue can determine whether a person has cancer.）
	问：有用动物做医学实验帮助治愈人类疾病的案例吗？**(Has animal testing led to cures for any human diseases?)** 答：有很多，动物测试提供了很多治愈人类疾病的医疗手段。比如：在狗身上做的实验帮助研发了治疗糖尿病的方案，在猴子身上做的实验帮助研发了治疗肝炎、小儿麻痹症、和艾滋病的方法。*(*Animal testing has led to treatments and cures for many human diseases. For example, research with dogs led to treatments for diabetes, research with armadillos led	问：可以用刚刚死亡的人体做医学研究吗？（**Can bodies of humans who have recently died be used for research?**） 答：可以，研究刚刚死亡的人体可以查明死因并且帮助了解药品的疗效，以便改进和研发药品。（Examining human bodies soon after death can help to better understand causes and effects of diseases and medicines.）

	to leprosy vaccines, and research with monkeys have led to treatments for hepatitis, polio, and AIDS.)	
Weaken opposing side (O-)	**问：有只能在动物身上进行而不能在人体身上进行的医学研究吗？(Are there types of research that can be performed with animals but not humans?)** 答：许多针对活体的医学研究非常复杂，其结果也非常不确定，因此这些实验只能在动物身上进行。比如，基因工程中针对如何改造动物器官以便在将来把它们移植到人体内的研究只能先在动物身上进行。(Many studies of living bodies are so complicated and uncertain in their effects that they could not be carried out with humans.)	**问：通过了动物实验的药品在人身上都应用成功了吗？**（Do most of the drugs that pass animal tests succeed in humans?） 答：美国食品和药品部门报道每100种通过了动物实验的药品中，有92种在人体身上应用失败。（The Food and Drug Administration reports that 92 out of every 100 drugs that pass animal tests fail in humans. ）
	问：除了医学实验，动物是否被用于其他原因？(Are animals used for other reasons than testing medical treatments?) 答：动物还可能被用于测试人体对新的化妆品或其他产品的反应。(Animals may be used to test reactions to new cosmetics or other products for the human body.)	**问：在实验室中的动物是如何被对待的？（How are animals treated in research laboratories?）** 答：尽管法律规定要把实验室的动物的痛苦和疼痛水平保持在最低水平，但是因为实验场所无法每分每秒都被监控，且实验记录是机密文件，所以现实中动物在实验室里是怎么样被对待的仍是一个未知数。（There are laws in place to help ensure that distress and pain in

		animals is kept to a minimum. However, the daily treatment of animals is not known because the testing places cannot be monitored at all times and records are not shared.）
Support opposing side (O+)	问：可以用人体细胞模型来测试新药吗？（**Can information be gained from studying human cells in a laboratory?**） 答：最新开发的由人体细胞制成的模型旨在更接近地模拟人体的工作方式，并可用于测试人类对药物的反应。（Studies of human cells under a microscope provide valuable information and can avoid the need for surgery. For example, examining human tissue can determine whether a person has cancer.）	问：为什么会用动物来做医学研究？**(Why have animals been used in research?)** 答：动物器官和人类器官常常非常相似，所以医药的疗效也通常很类似。(Animal organs often resemble human organs, so medicines may work in similar ways.)
	问：可以用刚刚死亡的人体做医学研究吗？（**Can bodies of humans who have recently died be used for research?**） 答：可以，研究刚刚死亡的人体可以查明死因并且帮助了解药品的疗效，以便改进和研发药品。（Examining human bodies soon after death can help to better understand causes and effects of diseases and medicines.）	问：有用动物做医学实验帮助治愈人类疾病的案例吗？**(Has animal testing led to cures for any human diseases?)** 答：有很多，动物测试提供了很多治愈人类疾病的医疗手段。比如：在狗身上做的实验帮助研发了治疗糖尿病的方案，在猴子身上做的实验帮助研发了治疗肝炎、小儿麻痹症、和艾滋病的方法。*(*Animal testing has led to treatments and cures for many human diseases. For example, research with dogs

		led to treatments for diabetes, research with armadillos led to leprosy vaccines, and research with monkeys have led to treatments for hepatitis, polio, and AIDS.)
Weaken own side (M-)	**问：通过了动物实验的药品在人身上都应用成功了吗？**（Do most of the drugs that pass animal tests succeed in humans?） 答：美国食品和药品部门报道每100种通过了动物实验的药品中，有92种在人体身上应用失败。（The Food and Drug Administration reports that 92 out of every 100 drugs that pass animal tests fail in humans.）	**问：有只能在动物身上进行而不能在人体身上进行的医学研究吗？(Are there types of research that can be performed with animals but not humans?)** 答：许多针对活体的医学研究非常复杂，其结果也非常不确定，因此这些实验只能在动物身上进行。比如，基因工程中针对如何改造动物器官以便在将来把它们移植到人体内的研究只能先在动物身上进行。(Many studies of living bodies are so complicated and uncertain in their effects that they could not be carried out with humans.)
	问：在实验室中的动物是如何被对待的？（How are animals treated in research laboratories?） 答：尽管法律规定要把实验室的动物的痛苦和疼痛水平保持在最低水平，但是因为实验场所无法每分每秒都被监控，且实验记录是机密文件，所以现实中动物在实验室里是怎么样被对待的仍是一个未知数。（There are	**问：除了医学实验，动物是否被用于其他原因？(Are animals used for other reasons than testing medical treatments?)** 答：动物还可能被用于测试人体对新的化妆品或其他产品的反应。(Animals may be used to test reactions to new cosmetics or other products for the human body.)

	laws in place to help ensure that distress and pain in animals is kept to a min-imum. However, the daily treatment of animals is not known because the testing places cannot be monitored at all times and records are not shared.）	

Appendix D

List of Q&A for Soda Topic Essay

1. **喝汽水在短期内如何影响大脑？(How does soda affect the brain in the short term?)**

当一个人摄入汽水中的糖分时，大脑中的奖励系统就会被激活。大脑会释放一种叫做多巴胺的化学物质，它会让人产生愉悦的感觉。(When a person consumes sugar, as found in soda, the reward system in the brain is activated. The brain releases chemical called dopamine, which produces feelings of pleasure.)

2. **喝汽水对健康有好处吗？(Does drinking soda have any health benefits?)**

一项针对有便秘症状的人的研究发现，喝碳酸水的人比喝普通自来水的人的便秘症状更少，而汽水中含有碳酸水. (A study of people with constipation symptoms found that those who drank carbonated water, as found in soda, had fewer symptoms than those who drank normal tap water.)

3. **所有类型的汽水都含有很多糖和卡路里吗？(Are all types of sodas high in sugar and calories?)**

现在有许多无糖或0卡路里的汽水。(Most sodas are available in no-calorie types.)

4. **关于汽水和肥胖的研究都是谁做的？(Who conducts most of the research on soda and obesity?)**

大多数发现汽水和肥胖之间存在联系的研究都是由反对饮用汽水的研究人员进行的。(Most studies that have found links between soda and obesity have been conducted by researchers who are against consuming soda.)

5. **不喝汽水可以改善健康吗？(Can avoiding soda improve health?)**

不喝汽水可以通过降低患高血压、哮喘、胰腺癌和糖尿病的几率来改善健康。(Avoiding soda can improve health by lowering the chances of hypertension, asthma, pancreatic cancer, and diabetes.)

6. **避免喝汽水可以帮助人们减肥吗？(Can avoiding soda help people to lose weight?)**

一项研究表明，减少汽水摄入量的高中生体重保持不变。而没有改变的学生体重增加了。(A study showed that high schoolers who reduced their soda intake maintained the same weight. Students who made no change gained weight.)

7. **汽水中的咖啡因所提供的短期能量提升能持续多久？(How long-lasting is the short-term energy boost from the caffeine found in soda?)**

研究发现，摄入咖啡因的学生更容易晚上睡不着觉，从而早上起来更容易感到疲倦。(Research has found that students who consume caffeine are more likely to have trouble sleeping and to feel tired in the morning.)

8. **喝普通汽水会导致体重增加吗？(Can drinking regular soda lead to weight gain?)**

研究表明，喝汽水可能会让人想吃得更多，因为它会导致血糖升高。 在一项研究中，在日常饮食中添加汽水的人多摄入了 17% 的卡路里。 摄入更多卡路里会导致体重增加。(Research suggests that drinking soda may make people want to eat more because it causes a rise in blood sugar. In one study, people who added soda to their regular diet consumed 17% more calories. Consuming more calories leads to weight gain.)

9. **喝汽水会让人感觉更有精神吗？(Does drinking soda boost energy?)**

大多数汽水都含有咖啡因，可以提供短期的能量供给，让人感觉更有精神和能量。(Most sodas contain caffeine, which gives a short-term energy boost.)

10. 喝汽水会影响牙齿健康吗？(Does drinking soda affect dental health?)

经常喝汽水会导致牙菌斑堆积在牙齿上，并可能导致蛀牙和牙龈疾病。(Drinking soda regularly causes plaque to build up on the teeth and can lead to cavities and gum disease.)

11. 学校允许学生在学校喝汽水吗？(Do schools allow students to drink soda at school?)

英国和法国的中小学都禁止销售汽水，中国的许多学校也正在考虑或已经禁止学校销售汽水。(In Britain and France, soda sales have been banned from elementary and high schools, and many US school districts are considering or already doing the same.)

12. 无卡路里汽水会影响一个人的体重吗？(Do no-calorie sodas affect a person's weight?)

无卡路里汽水用甜味剂（这是一种糖的替代品，有甜味但不是糖）来替代真正的糖。研究人员发现甜味剂对体重几乎没有影响。(Researchers claim that sugar substitutes, such as those found in no-calorie sodas, have little to no impact on weight.)

Appendix E

List of Q&A for Animal Research Topic Essay

1.为什么会用动物来做医学研究？(Why have animals been used in research?)

动物器官和人类器官常常非常相似，所以医药的疗效也通常很类似。(Animal organs often resemble human organs, so medicines may work in similar ways.)

2.有用动物做医学实验帮助治愈人类疾病的案例吗？(Has animal testing led to cures for any human diseases?)

有很多，动物测试提供了很多治愈人类疾病的医疗手段。比如：在狗身上做的实验帮助研发了治疗糖尿病的方案，在猴子身上做的实验帮助研发了治疗肝炎、小儿麻痹症、和艾滋病的方法。*(*Animal testing has led to treatments and cures for many human diseases. For example, research with dogs led to treatments for diabetes, research with armadillos led to leprosy vaccines, and research with monkeys have led to treatments for hepatitis, polio, and AIDS.)

3.可以用人体细胞模型来测试新药吗？（Can information be gained from studying human cells in a laboratory?）

最新开发的由人体细胞制成的模型旨在更接近地模拟人体的工作方式，并可用于测试人类对药物的反应。（Studies of human cells under a microscope provide valuable information and can avoid the need for surgery. For example, examining human tissue can determine whether a person has cancer.）

4.可以用刚刚死亡的人体做医学研究吗？（Can bodies of humans who have recently died be used for research?）

可以，研究刚刚死亡的人体可以查明死因并且帮助了解药品的疗效，以便改进和研发药品。（Examining human bodies soon after death can help to better understand causes and effects of diseases and medicines.）

5.动物和人类得的疾病类似吗？(How similar are humans and animals in terms of diseases they get?)

许多人类的疾病，例如癌症、疟疾、哮喘、关节炎和心脏病，也存在于动物身上。(Many of the diseases that humans get—such as cancer, malaria, asthma, arthritis, and heart failure—are also found in animals.)

6.有只能在动物身上进行而不能在人体身上进行的医学研究吗？(Are there types of research that can be performed with animals but not humans?)

许多针对活体的医学研究非常复杂，其结果也非常不确定，因此这些实验只能在动物身上进行。比如，基因工程中针对如何改造动物器官以便在将来把它们移植到人体内的研究只能先在动物身上进行。(Many studies of living bodies are so complicated and uncertain in their effects that they could not be carried out with humans.)

7. 通过了动物实验的药品在人身上都应用成功了吗？（Do most of the drugs that pass animal tests succeed in humans?）

美国食品和药品部门报道每100种通过了动物实验的药品中，有92种在人体身上应用失败。（The Food and Drug Administration reports that 92 out of every 100 drugs that pass animal tests fail in humans.）

8. 在实验室中的动物是如何被对待的？（How are animals treated in research laboratories?）

尽管法律规定要把实验室的动物的痛苦和疼痛水平保持在最低水平，但是因为实验场所无法每分每秒都被监控，且实验记录是机密文件，所以现实中动物在实验室里是怎么样被对待的仍是一个未知数。（There are laws in place to help ensure that distress and pain in animals is kept to a min- imum. However, the daily treatment of animals is not known because the testing places cannot be monitored at all times and records are not shared.）

9. 问：医学研究者可以在实验室中无止境地滥用动物吗？(Can researchers use as many animals as they wish in their research?)

医学研究制度规定在实验中要用最少量的动物来达到研究目的。(Regulations exist that require that scientists use as few animals as possible to conduct their research.)

10. 问: 美国每年有多少只动物被用来进行医学研究？(How many animals are involved in research each year in the US?)

据美国农业卫生部报道，1990年代初期科研人员在实验中使用的动物数量达到高峰，每年有超过200万只动物被用于实验研究中。之后这一数字逐年下降，2016年被用于做实验的动物数量下降至100万只以下。（The US Department of Agriculture reports that use of animals in research was at its highest of over two million per year in the early 1990s and fell to a low of below a million in 2016.）

11. 问：实验室的动物参与了实验以后的结局是什么？（What happened to animals after they participated in laboratory research?）

有些动物继续参与后续的实验，有些动物在实验的过程中死亡，有些动物在实验后被注射安乐死的药物以避免疼痛。因为实验室的动物大多是为实验专门培育的，因此它们不适合被放到野外独立生存。(Some animals continued to participate in subsequent experiments, some animals died during the experiment, and some animals were euthanized after the experiment to avoid pain. Because most laboratory animals are specially bred for experiments, they are not suitable to be released into the wild to survive independently.)

12. 问：可以用数据分析来研究人类对不同事件的反应吗？（Can statistics be used to analyze how people react to different events?）

统计学家通过对人类行为的大数据分析发现了吸烟和肺癌、饮食健康和心脏病之间存在联系。（Statisticians have helped link cigarette smoke to lung cancer and diet to heart disease by studying large numbers of people over periods of time.）

Appendix F

Constructed Dialogue Task

In most western countries, such as the United States, Teens who commit serious crimes may have to appear before a judge in an adult court system. However, some people disagree and think that teens are better served by a court just for juveniles than by a regular adult court.

*Li Ping and Wang Hua are two expert arguers and well matched. Now they are arguing about whether teens should be tried in an adult court or juvenile court. Write a script of what they might say. Your script should present the best arguments you can construct.

Begin your script like this:
Li Ping: I think teens who commit serious crimes should be judged in a juvenile court because juvenile court is designed specifically for teens.
Wang Hua: I disagree. Even though it is designed specifically for teens but it doesn't mean that it is working effectively.
Continue the script, filling in what each one might say:
Li Ping: XXXXX
Wang Hua: XXXXX

Here is some information on the topic. You might find it helpful, but you are not required to use it in your script.

1. The judges and staff in a juvenile system are specially trained to deal with young people in trouble. Punishments tend to be less severe and sentences shorter in juvenile court.
2. A "get tough" policy has become more popular in recent years, with a federal law proposing that adolescents as young as 16 are tried in regular adult court.
3. The prefrontal cortex, which is responsible for abstract thinking and the ability to exercise good judgment, is not fully developed until one's early- to mid-20s.
4. Government would save money if they didn't have to pay for a separate juvenile system. Juvenile courts and prisons need more people to run and thus cost more. Adult courts cost less to operate.
5. Teens commit violent crimes such as murder. Teens were involved in one quarter (25%) of violent crimes in the USA over a 25-year period beginning in 1990.
6. They do not get records if sentences are served in a juvenile detention center; their records are sealed on their release.
7. In the US, the rate of recidivism (repeat crime) ranges from 68% for those under age 21 to 16% among those over age 60.
8. Teens are at risk of being assaulted in adult prisons. 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.