Assessing Argumentation Skills

Karen Sybille Bruun

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

Assessing Argumentation Skills

Karen Sybille Bruun

Skills of argument have attracted the attention of educators but remain challenging to both assess and develop. In contrast to the traditional essay, dialogic argument requires reflection on and coordination of one’s own claims with those of an interlocutor. Investigating a tool for assessing an individual’s dialogic argument skill is an objective of the present work. Building on an earlier study by the author and colleagues, and informed by philosophical writings on objectives of argumentation, undertaken here is a conceptual analysis of instances of dialogic argumentation by skilled arguers in order to discern its essential characteristics. The identified set of characteristics is then used as a basis for evaluating the argumentation skills exhibited by a sample of sixth grade students. A practical purpose is development of an assessment tool for use in educational contexts, identifying the range and variation of argumentation skills individuals bring to dialog. A value of the individual instrument referred to as a constructed dialog and developed and employed here, is that it overcomes the statistical problem created by lack of independence between participants in a dialog which requires that the unit of analysis be the pair- thereby defeating the objective of assessing the skill of an individual.

Empirical results document that young adolescents display competence in some basic skills of argumentation but, even following an intervention designed to build and exercise such skills, they continue to use these sparingly and to lack other equally fundamental ones.
Discussion addresses implications for education, as well as the potential for use of the constructed dialog as an assessment tool for evaluating an individual student’s skill in argumentation and the associated understanding it reflects regarding the nature and objectives of argumentation.
# Table of Contents

List of Tables and Figures........................................................................................................... ii

Acknowledgments.................................................................................................................... iii

Dedication .................................................................................................................................. v

Chapter 1: Introduction.............................................................................................................. 1

Chapter 2: Literature Review .................................................................................................. 7

Chapter 3: Methods ................................................................................................................ 15

Chapter 4: Results .................................................................................................................. 34

Chapter 5: Discussion ............................................................................................................. 50

References .............................................................................................................................. 59

Appendix ................................................................................................................................. 68
List of Tables and Figures

Table 1  Summary and Examples of Coding Scheme for Constructed Dialogs…… 28
Table 2  Examples of Successful and Unsuccessful However Unit-pairs………. 31
Table 3  An Example of a Constructed Dialog from the Present Study………. 32
Table 4  Constructed Dialogs Intervention Group/ Pretest………………………… 35
Table 5  A-only Group/ Pretest (Xiao & Kuhn, 2024)………………………… 35
Figure 1  Mean Idea Unit (IU) frequencies in Constructed Dialogs by Condition and Time…………………………………………………………………… 38
Figure 2  Mean Counter-Alternative (CA) frequencies in Constructed Dialogs by Condition and Time…………………………………………………………………… 40
Figure 3  Mean Counter-Critique (CC) frequencies in Constructed Dialogs by Condition and Time…………………………………………………………………… 41
Figure 4  Mean Counter-Undermine (CU) frequencies in Constructed Dialogs by Condition and Time…………………………………………………………………… 42
Figure 5  Mean Evidence (EV) frequencies in Constructed Dialogs by Condition and Time…………………………………………………………………… 43
Figure 6  Mean However Statements (HS) frequencies in Constructed Dialogs by Condition and Time…………………………………………………………………… 44
Table 6-1  Intervention Group Summary (Present study) …………………… 46
Table 6-2  Xiao & Kuhn (2024) Comparison…………………………………… 47
Table 7  Estimation Results of Negative Binomial/ Poisson Regressions on Post-intervention frequencies………………………………………………………… 49
Acknowledgments

Dr. Deanna Kuhn, my academic advisor - you granted me access to a world that I had always hoped to explore. With a combination of incredible generosity and gentle guidance, you have shepherded me through this process. You have always gone above and beyond in your support, and I can’t thank you enough for facilitating this incredible experience.

I would like to express my deepest gratitude to my friend and mentor, Dr. Ben Lovett. There is a world in which our paths never cross and I am certain that, in that world, I would be wandering the streets of Upper Manhattan, gazing wistfully at the stone buildings of TC, wishing I were inside. You have guided me, unwaveringly offered your intellectual wisdom and insight and, thankfully, remain my dear friend.

I would like to express my deepest appreciation to my thesis committee for your time and efforts- Dr. Black, Dr. Rubin, and Dr. Metcalfe.

I would be remiss if I didn’t mention DK’s lab members who, over the years, have become colleagues and friends. Teresa Fraguda- you and I have been on this journey for quite some time. You have been a sounding board and a thoughtful collaborator, and I thank you for all your time and efforts as we have made this journey. Si Xiao, Yu Song, and Mariel Halpern- you forged the path and made it look effortless. Thank you for your guidance, friendship and support.

Dominic A.A. Randolph – you believed in me and gave me the space in which to pursue my graduate work. I am deeply thankful.
It took a village to help raise my family while I completed this dream:

To each of my dear friends - Whitney Busch, Maxon Davis, Emily Edwards, Hannah Lavan, Kara Melley, Davia Moss, Ellen Ochu, Sara Ston and Kenley Weaver- your care and your friendship have been truly invaluable to me and helped me get to the finish line. Thank you for being you and for enriching my life.

Lydia, David (and Georgie) Federman - you helped every step of the way, whether convenient (or otherwise) and this would never have happened without your support. We love you like the family you are.

Julia Guerrero - you raised the boys when I couldn’t, and we are so lucky to have you as our Tita. You are my second mother and I am so thankful for all your love and wisdom.

My sister, Sophie Bruun and my brother-in-law, Benjamin Olivares - you loved, listened, cared and encouraged. There is no me without you.

To my boys, Isidore and Jude - you will forever teach me more than I can hope to teach you and you have given me purpose for life. May this dissertation always be a reminder to you that there is nothing you can’t achieve if your heart is in it.

Lastly, to my husband, Tyler Moss - you are my home and your love was the place from which I could dream this dream. Thank you for making my dreams come true.
Dedication

This thesis is dedicated to my father, Hans Henrik Bruun.

You gave me a love of reading and, thereby, opened the whole world to me.

I am forever indebted to you.
Chapter 1: Introduction

The art of critical thinking is now, more than ever, a necessary skill for students to develop and its benefits are both extensive and lauded (Best, 2021; Ennis, 1991; Hatcher, 2005; Sternberg, 2006) According to Walton, “the common core of basic critical thinking skills underlying critical reasoning in each discipline is the key ability to look at both sides of an argument” (Walton, 1989, p.182) In the current political environment, this ability to look at both sides of an argument is waning and schools are increasingly forced to respond to a climate of national enmity as political divisions of historic proportions make daily headlines. Nowhere is this debate more salient than in the schools and colleges where both the merits and challenges of campus discourse are hotly debated. And while the ethical and moral implications of this campus discourse is under constant scrutiny, less examined is how to evaluate the quality and nature of that discourse.

In schools, discourse opportunities are often linked to activities rooted in inquiry and scientific learning. Debate, as well as individual essay construction and writing, occur in this inquiry and argument in practice - as purposeful peer-to-peer exchanges (Kuhn, 2019). Building on a social constructivist, as well as a socio-cognitive framework, a dialogic method is practiced in these classroom interactions (student-teacher/ student-student inquiry and argument), where students are iteratively building upon knowledge as it develops as well as being challenged by others to develop new ideas and new ways of thinking (Lyle, 2008; Murphy et al., 2018). Advocates of dialogic methods have emphasized peer interaction as a method to develop individual argument writing skills, as well as dialogic skills themselves (Kuhn & Crowell, 2011; Kuhn et al., 2013; Shi et al., 2019). In the literature review presented in Chapter 2, summarized
is the long-standing theoretical literature regarding the merits of a dialogic method and the now extensive empirical research literature in psychology and education that documents its outcomes.

Peer dialog and the use of argumentation are now widely endorsed as educational methods across age levels and disciplines (Kuhn et al., 2022; Littleton & Mercer, 2013; Lyle, 2008; Resnick et al., 2015, Wolf, et al., 2006) However, measuring skill in individual or dialogic argument is a difficult task and recent scholarship suggests that the construction of tools with which to measure the quality of individual argumentative writing has been elusive (Hahn, 2020) and at best, incomplete. Even less work has been devoted to developing tools for assessing dialogic argument skill, which is the main objective of the present dissertation. If dialogic argumentation skill is both an important skill in its own right and a path to the development of skill in individual written argument, it is essential to have in place a tool that assesses this argumentation skill.

The present study is part of a two-pronged effort undertaken in the past several years by a research group at Teachers College, Columbia University led by Deanna Kuhn. The first prong has consisted of conceptual analyses of instances of skilled dialogic argumentation to discern its essential characteristics. This work has been informed by theoretical writings of philosophers on the nature and objectives of argumentation, as well as samples newly collected and analyzed by the present author. The second prong has consisted of development of an assessment tool for use in educational contexts, for the purpose of identifying the argumentation skills an individual brings to a dialogic context and their variation across individuals, as this variation would be expected to play a role in the success of dialogic encounters in which an individual participates.

The present study contributes to both of these prongs. It first of all builds on a previous study (Kuhn et al., 2024) in which individual participants (community adults) were asked for
their views on a social issue (DACA-Deferred Action for Childhood Arrivals), after viewing one of two videos. In a monologue condition, the video portrayed two skilled actors individually presenting their opposing views (send them back vs. let them stay) on the immigration topic; in a dialog condition the two actors engaged in a respectful exchange of ideas, addressing one another’s claims but expressing the same ideas as they presented in their individual presentations. Participants were then asked to express their own views and to answer an interviewer’s follow-up questions. These responses were subsequently coded on multiple dimensions of cognitive complexity. Findings indicated that exposure to dialogic discourse yielded richer responses, compared to the monologic condition.

The dialogic and monologic videos viewed by participants in this study appear in Appendix 1, as they demonstrate differences across conditions of a kind that were later observed in participants’ own expressed views on the topic. The ideas expressed in the two videos are seen to be identical with only minimal elaboration in the monologic video (so that the two videos remained the same length), but in the dialogic video the speakers reveal several dimensions of skilled argumentation (see Appendix 1). In short, they respond to what each other has said, engaging, questioning, and critiquing it, as well as commenting on the dialog itself. These characteristics paralleled characteristics that were then observed to a greater extent in the responses of participants who viewed the dialogic exchange, compared to the group who viewed a monologic video in which the two actors individually and sequentially expressed the same ideas.

These characteristics are consistent with those identified by philosophers of argument (Macagno & Walton, 2015; Walton 1989) regarding the objectives of skilled dialog: to question the opponent’s position by identifying and challenging gaps in their argument, and to secure
commitments from the opponent that can be employed to advantage in the exchange. The characteristics validated in this study as characteristics of skilled argumentation can thus be drawn on as a basis for establishing a tool to assess the level of skill observed in an individual’s argumentation – a second goal of the present work.

The immediate challenge in developing such a tool is that argumentation is dialogic, involving two or more participants, and our goal here is to assess an individual’s skill in such argumentation, absent the variation introduced by the nature of the dialogic partner’s input. Assessing dialogic argumentation is also challenged by the statistical problem that this lack of independence between the two participants requires that the unit of analysis be the pair, which defeats our purpose of assessing the skill of an individual. For this purpose, we have developed a task we refer to as a “constructed dialog,” introduced by Zavala and Kuhn (2017) with adults and subsequently revised by our current research group in several iterations for use as an assessment tool. As the second part of the present study, we examine constructed dialogs as an argumentation skill assessment tool, based on the study of a sample of 6th grade students at an academically high-performing school.

The constructed dialog task, as it was used in the present study, asks an individual student to construct in writing a dialog between two imaginary interlocutors who hold opposing views on a topic. They are described as skilled arguers and evenly matched. A constructed dialog makes it possible to assess the argumentation skill of an individual, unconfounded by the influence and level of skill of an external interlocutor. In addition to the proposed merit of constructed dialogs as sensitive tools in their ability to assess individual skill in argumentation, they also have the potential to reveal an epistemological level of understanding regarding what a dialogic argument consists of. The use of constructed dialogs as a factor in shifting epistemological beliefs from
basic to more sophisticated has also been explored among adults (Zavala and Kuhn, 2017) with results suggesting that performance on the constructed dialog task was related to participants’ levels of epistemological understanding.

**The Present Study**

The broad purpose of the present study is to develop and apply an instrument that will be of value for assessing the argumentation skill of individual middle-school students, an age at which argumentation begins to play a role in the curriculum and becomes increasingly important to academic progress, as well as being the age range when most educational research on argument has been conducted (see Chapter 2 and Rapanta & Felton, 2021, for review), with this age group in general showing weak argument skills in need of further development. In addition to Zavala’s research with adults, a study by Shi (2019) in China with younger students employed a similar instrument but only as one of several outcome measures following an intervention. We therefore wished to examine the constructed dialog instrument as a tool both prior to and following an intervention known to support argument skill development.

Three-quarters of 8th and 12th graders in the US lack proficiency in writing (NAEP, 2008) We chose to study an educationally advantaged sample of students, attending a school where argumentation and debate are already featured in the curriculum. Would this educational advantage establish that it had in fact led these students to achieve a high or even maximum level of individual skill in argumentation, or would a known successful intervention elicit further gains? Some academically lower achieving students are also members of this school’s student body and enhanced by measurement both before and after an argument-targeted intervention, we predicted their inclusion would yield a wide range of student skill levels. To ensure this purpose would be met, we compare the findings of the present study to those of another member of our
research group whose study (Xiao & Kuhn, 2024) was conducted contemporaneously with the present one and included the constructed dialog task used in the present work as part of its design. The similarly aged population, however, was from a school whose curriculum included little to no experience in argumentation or argument writing.

In sum, the present design was thus expected to reveal potentially more advanced skill levels, as well as less skilled ones, that would make skill-level distinctions clearer and of wider range, thereby enriching evaluation of the potential of constructed dialogs as an assessment tool in measuring students’ skill in argumentation in both research and applied settings. Doing so sets the stage for comparison with other assessment tools, both potential and traditional. If, as advocates of dialogic approaches maintain (Rapanta & Felton, 2021), dialogic argumentation provides a pathway to advances in traditional argumentive essay skills critical to academic success, then identifying and assessing the skills entailed in dialogic argumentation is an important effort.

---

1 *Argumentive* has been previously used by Kuhn and her colleagues to express the disposition and function of argumentation and has become widely adopted in the literature in recent years. This differs from *argumentative*, a verb used to express a person’s ability to perform argumentation.
Chapter 2: Literature Review

“All human life could well be conceived as a great discourse” (Kierkegaard)

The exploration of cognitive collaboration is crucial in an educational context and a rich theoretical framework underpins the dialogic method. In this chapter, we will attempt to clarify the constructs which form the basis for the dialogic method as well as examine the empirical evidence contributing to this rich history.

Philosophically, dialog as a method of reasoning (dialectics) has its western roots in Greek antiquity as a way in which nuances could be explored during an exchange of ideas (Plato, 380 BC/1910) and, later, as a theoretical basis for an ethical existence (Bergman, 1991) or as the root of relationships, the way in which to “all experience one another” (Buber, 1967, p. 85). According to Walton, “the intended use of dialog theory was to provide a normative structure on which to ground methods for improving critical thinking skills, including writing skills” (Walton, 2014, p. 9). In this work, Walton chronicles how the developments in argumentation theory shifted from over-simplified judgements of theoretical fallacies to a more practical and systematized attempts to grapple with everyday arguments (van Eemeren & Grootendorst, 2003).

Developmentally, origins can be traced to Vygotsky’s theory of social development (1937/1987) detailing the shift from interiorization to dialog in the social context, occurring when a young child is faced with a type of change that interrupts a flow of activity. This causes an increase in egocentric speech and this language soon becomes a way children can solve
problems- “in seeking and planning the solution to a problem” (Vygotsky, 1962, p.16). Bakhtin would suggest that this language, in response to earlier expressions (implicit or explicit), constitutes discourse (Emerson, 1983, p. 248) and, with development, this discourse becomes social in nature. This development lays the foundation for the socially collaborative learning which underpins educational programming today.

The perspective adopted here is that exploring thinking in a social context allows us to explore more deeply why we hold the opinions we do and affords us the opportunity to consider discourse in a real-world context - thinking as argument (Kuhn, 1992). In other words, social discourse becomes a way to bridge external experiences, transforming them into internal reflections. In schools, these discourse opportunities are often linked to activities rooted in inquiry - scientific learning, debate and essay construction and writing.

**Argumentation as thinking**

The purpose of argument is not only to make one’s own position understood and known but also to effectively refute the position of the other (Walton, 1989). A definition of argument is therefore necessary to define the parameters within which the present study is situated. The type of argument under scrutiny is dialogic argument – defined by Kuhn (1992) as an exchange in which participants recognize two opposing assertions and, as part of an integrative evaluation, relate supporting and refuting evidence in service to one position. Implicit in this exchange of ideas is the structure within which the argument is taking place - the context of conversation in which there is communication and exchange which encompass the goal of persuading the other (Walton & Krabbe, 1995). As children develop argument skills, reasoned argument - the weighing of contrasting statements - becomes an inherently social activity and the construction of an argument becomes discursive in nature (Felton & Kuhn, 2001; Kuhn, 1992), reliant on the
responses of others (Kuhn, 2019). Dialogic argumentation in this present study has its roots in the conception of thinking as an intrinsically social endeavor and the dialogic approach is at the heart of programs like Kuhn’s “Argue With Me” (Kuhn et al., 2016a). We explore this program next, considering the developmental implications and empirical studies upon which the present study was based.

**Argue With Me**

Seeking an understanding of the role of argumentation in the development of higher-order thinking skills has prompted researchers to examine, and educators to implement, programs which claim to support the development of these skills (Asterhan & Schwartz, 2007; Asterhan & Schwartz, 2016; Nussbaum & Schraw, 2007). The “Argue With Me” (AWM) program has been presented and implemented both as intensive short-term programming (Xiao & Kuhn, 2024) as well as longer-term interventions (Crowell & Kuhn, 2014; Kuhn, et al. 2016a, 2016b), embedded in school curriculum, and has demonstrated significant empirical successes in the development of argumentation and content skills in adolescents in areas such as counterfactual reasoning, dual-perspective reasoning and the integration of opposing arguments and rebuttals (Iordanou, 2013; Kuhn & Crowell, 2011; Kuhn et al., 2008, 2013; Mayweg-Paus et al., 2015) as well as the use of evidence/ evidence-based segments (Hemberger et al., 2017; Iordanou & Constantinou, 2015), the transfer of argument skills through a collaborative writing intervention (Matos, 2021), the increase of evidence use (Shi, 2019), and perspective-taking (Hemberger et al., 2017; Iordanou et al., 2019; Kuhn & Crowell, 2011, 2014; for a full review see Iordanou & Rapanta, 2021). With the positive impact of well-documented peer interaction of cognitive development (Resnick, et al., 2015), the AWM program has been used in educational settings with small groups where teacher-coaches facilitate students as they consider a cognitively demanding topic.
(meaning a topic with no simple answer) and engage in peer-to-peer discourse in what is described by Kuhn et al. as “a community of practice”. AWM and the perspective-taking which students implicitly undertake as part of the program supports students to develop both the ability to weigh opposing positions and include negatives of an opposing position in argument— the dual perspective— as well as (as a higher-order thinking skill) the ability to include negative aspects of the arguer’s own perspective as well as positive aspects of the opposing position— the integrative perspective. The embedding of counterfactual reasoning within these perspectives is significant as students consider positions that are not their own and have to come to terms with evidence that supports them, as well as evidence that may not support their own positions (Kuhn & Crowell, 2011). These findings support Graff’s (2003) claim that by having students determine both what their own perspective is and what it isn’t, student writing reaps the benefits. The next area examined here is to what extent these positions can be “staked out” with no interlocutor and how the benefits of doing so can be reliably measured.

**How argumentation skill develops**

Understanding that targeted interventions can support children in developing argumentation skills, the question remains: do the skills necessary for argumentation develop naturally and, if so, at which point(s) of development? Significant research demonstrates that children develop argumentation skills naturally and can produce rudimentary arguments at an early age (Clark & Delia, 1976; Eisenberg & Garvey, 1981; Mercier, 2011; Stein & Bernas, 1999) as well as engage in debates (Mason, 2001). Equally, children reflect basic epistemic cognition already at pre-school age (Iordanou, 2017) and, as early as three years of age, will use evidence to support a claim (Köymen et al. 2014). Research suggests that epistemic cognition changes over time (Iordanou & Constantinou, 2015; Kuhn & Moore, 2015) and, as children
develop, so does their ability to distinguish causality from covariation (Kuhn, 2012) as well as the development of metacognitive monitoring (Kitsantas & Zimmerman, 2002; Roderer & Roebers, 2014). However, limitations to this development exist as children and teens often neglect to consider multivariable causality and instead focus mainly on a single cause as the reason for a given outcome (Felton & Kuhn, 2001; Kuhn et al., 2020; Pease & Wirkala, 2009). In addition, there is evidence from Kuhn’s 1992 study that less than a quarter of adults examined were classified as comfortable when thinking about, evaluating, and engaging in, argument. Instead, results indicated a need to see the point of an argument in order to engage with it. Kuhn calls this tendency to avoid critical aspects of argumentation “epistemological naivete” (Kuhn, 1992, p. 168). This lack of understanding of the basic value of argumentation lends indicates a need to implicitly teach the value of this skill to children, teens and, when they are willing, adults.

In order to promote a continued epistemological development, there is evidence to suggest the benefits of engaging in inquiry as well as practicing the process of argumentation impact change in epistemological beliefs (Iordanou, 2016) improving adolescents multivariable causal reasoning ability (Kuhn et al., 2017).

**Constructed Dialogs**

Building on this foundational work, researchers have further examined to what extent a dialogic argument written by an individual student, taking place between two hypothetical interlocutors with opposing views, might more comprehensively allow those students to demonstrate richer argumentative skills than traditional assessment types. The structure of the constructed dialog task encourages the weighing of arguments and counterarguments as participants naturally shift between opposing viewpoints. While it may have been the position in
the past that dialog had to be socially mediated in order to confer any benefit on a collaborative activity, this may not necessarily be the case. Inexperienced arguers in a peer-to-peer interaction may not be aware of whether a contribution to an argument is useful and moves the debate forward or what function it may serve. As they are a solo activity, constructed dialogs afford students the opportunity to actively consider this while risking little in the process. The use of constructed dialogs as a post-assessment, in conjunction and/or in comparison with traditional essays, has been explored by several researchers. In 2013, Kuhn, Zillmer, Crowell and Zavala examined to what extent the use of a constructed dialog task demonstrated an effective use of candidate statements as evidence to support the arguer's claim, addressing not only the strengths of the interlocutor's preferred candidate but also the relative weaknesses of the opponent. Results indicated that the intervention group demonstrated a deeper understanding of discourse norms. Additionally, the use of constructed dialogs as a factor in shifting epistemological beliefs from basic to more sophisticated was explored by Zavala in 2016, with results strongly suggesting that the constructed dialog task had a significant effect on the epistemological understanding of participants. Furthermore, Zavala and Kuhn (2017) explored the use of constructed dialogs to investigate the development of argumentative discourse norms, as well as effective levels of evidence use. This study revealed that the constructed dialogs composed by intervention participants included evidence in a more integrative manner than did essays of a comparison group, drawing on several aspects of presented evidence to make their opposing claims. The authors propose that requiring a writer to make a constant cognitive shift between positions support students to develop richer representations of the relationship which links one position to another position. In a 2015 study by Kuhn and Moore, 90% of middle-school students used more evidence from their own personal experiences in their dialogs than in their essays, where they
used only 40%. These results suggest that the engagement in their dialogs was richer and felt more personal. Even the observation of dialogs between partners who engage in civil discourse with opposing views seems to have positive effects, despite no active engagement on the part of the observer (Chi et al., 2016; Kuhn, et al., 2024). Constructed dialogs appear to have potential as an assessment in and of itself, with researchers more clearly being able to determine growth at the individual level with greater statistical accuracy.

In order to determine to what extent students successfully exhibit rich argumentation skills, the development of an evaluative system for argumentative writing has been necessary. Over time, a coding system drawing on Walton’s (2005) coding scheme has developed, focusing first on the use of argument/counter-argument. Traditionally, however, peer-to-peer dialogs have not served the function of assessing individual skills, as there is inherently a dualistic nature to a shared written experience between peers. Therefore, without the influence of an interlocutor, constructed dialogs with an imagined other have the methodological advantage of assessing the argument skills of an individual, since the unit of analysis is no longer shared.

**Present Study**

Previous research, serving as a foundation for the present study, has demonstrated proven gains in argumentation skills for students due to the AWM intervention. However, little work has been done in highly rigorous educational settings where students already have debate and discourse opportunities as a regular part of a set curriculum. Therefore, the impact of the AWM program in such a setting has yet to be examined. In addition, while the use of constructed dialogs has been established as an educational tool to support students in developing these skills, few studies have examined constructed dialogs as a stand-alone assessment tool to measure individual students’ skill in argumentation. Therefore, the present study aims to examine the
impact of AWM in a rigorous setting with the primary research question being how the constructed dialog assessment tool may reveal skill-level distinctions and how these distinctions may be related to participants’ levels of epistemological understanding of argumentation.
Chapter 3: Methods

Participants

Participants were 74 sixth graders, aged 11-0 to 11-9 years, attending a selective-admission, high-achieving urban independent Pre-K-12 school located in the greater New York area. Of the 74, 50% identified as female and 50% identified as male, reflecting the roughly equal gender distribution of the school population. The overall demographic of the school includes 44% students of color and 20% students receiving financial aid. The school serves a socioeconomically privileged student population with approximately 30% of graduates matriculating to Ivy League colleges and 100% of students accepted to four-year colleges upon graduation. Participation was voluntary and participants were able to opt out at any time. Four students left the workshop due to scheduling conflicts after the workshop commenced.

Design

With the approval of the school’s headmaster and other departmental heads and administrators, all sixth-grade families at the beginning of the 2022 school year (n=87) were offered the opportunity for their students to participate in a 12-session workshop, described to them as an optional Fall debate and dialog skill development workshop, promising to strengthen the critical thinking skills of all students, where participants would gain practice in the academic skill of taking and justifying positions on an issue both individually, verbally or in writing, and in collaborative debate with peers. The expectation was that students in the intervention group would, through the experiences and skills gained in the workshop, demonstrate superior
argumentation skills following the intervention, relative to the non-intervention comparison group.

A total of 52 families accepted the invitation and their sixth graders began the workshop; of these, 39 completed the workshop, as well as the prerequisite pre and post assessments, and constitute the experimental group. The remaining 35 sixth-grade students constituted a comparison group; they did not participate in an intervention but did participate in all pre- and post-test assessments. To achieve optimal group size and to allow participation by all interested students, the 52 students in the experimental group were divided randomly into two groups, the first of 29 students and the second of 19 students. The first group began meeting early in the school year, and the second group began meeting shortly after the first group had completed the workshop.

**Pre- and Post-intervention Assessments**

Students in both the intervention and non-intervention comparison group completed a pre- and post-intervention assessment in their English classrooms, designed to assess their individual skill in argumentation – the constructed dialog task introduced in Chapter 1. The pre-intervention task took place one week prior to the beginning of the workshop. The post-intervention was identical and took place three weeks after intervention students completed the workshop. These assessments were administered by the homeroom English teachers (three teachers in total) who had all been instructed in giving the assessments. Both assessments were completed online in a Google Forms format to ensure uniform experiences for students across different teachers. The researchers were not present during these assessments and students were not told that the assessments had any connection to the workshop/ intervention.
The topic addressed in the pre- and post-intervention assessments was not addressed in the workshop/ intervention. The topic was unrelated to any class/ curriculum work, never came up in the workshop and was framed as a formative pre-assessment for the English classroom. The classroom teacher let the students know that these were formative assessments that would help teachers prepare future lessons and therefore students were encouraged to try their best and take the assessments seriously. The prompt for the constructed dialog was stated as follows, both in the Google Form and on the physical paper copy presented to students. All students were asked to read the hard paper copy before they began the online writing:

“Lex and Robyn have different positions. Lex favors adult court for teens who commit serious crimes. Robyn favors juvenile court. They are discussing the topic. They are both expert arguers and evenly matched. Write a script of what they might say.

Begin your script in the space below like this:

LEX: I think they should be sent to regular adult prison because...(complete)

ROBYN: I disagree because...

Continue in the same way, filling in what each might say.”

Additionally, the students were given a sheet of paper (see Appendix 2) with definitions for each of the words in the prompt that may have appeared unfamiliar, such as juvenile and juvenile court. All but two students had completed the task within a 15-min time period. Two students asked for five additional minutes and were allowed that time.
Intervention content and procedure

The argumentation intervention was one known to be effective in developing argument skills in young teens. It followed the rationale and procedures developed by Kuhn and colleagues (2018a, 2018b; Kuhn, Hemberger & Khait, 2016; Hemberger et al., 2017). However, its length was much abbreviated due to limitations posed by the school on how much of the sixth-grade school schedule could be devoted to the activity. The agreed-upon 12 sessions (meeting twice in an 8-day cycle except when other activities reduced meetings to once a cycle) that met during the school term, however, we judged would be sufficient to accomplish our purpose of engaging them in argumentation in a format that would maximize the skill levels they displayed.

A team of researchers from Teachers College implemented the in-person intervention over 12 class periods, each period 50 minutes in length. The class met at rotating times of day (in accordance with the school schedule) twice weekly except when vacations or other school activities intervened, over a period of 16 weeks. The class was known as “Talk Time” (TT). The intervention was introduced to the students as a dialog and debate skill-building class and the students participated willingly.

Classroom setting. The first workshop meeting took place during the third week of the school year, allowing sixth graders to first acclimatize to the new school year. Three experienced adults led the workshop (one of them the present author). Each of the three had experience with the Argue With Me/TT method and had previously participated in at least one workshop using this method. Three student assistants also participated at various times, to provide logistical support, but had no prior experience. Assistants did not provide any direct instruction and mainly circulated to help the students stay on task or manage physical materials. All assistants were blind to the research questions.
**Topic.** A debate topic was chosen for the intervention by the researchers in consultation with the workshop teachers, prior to the beginning of the workshop. It was chosen as a topic with no ‘right answer’ where rich and nuanced arguments could be made on both sides of the issue. It was also deemed necessary to ensure that the topic was engaging but not overly polarizing and appropriate for the age range involved. It also needed to be a topic that could transcend socio-economic barriers, did not require any extensive pre-existing knowledge, and was one that this age group found relevant and felt entitled to have an opinion on, thereby engendering engagement and enthusiasm in debating. Meeting these requirements, the question chosen was this one:

*You are very good in one school subject and don’t do well in another. Should you put most time and effort in being at the top of the class in your strong subject or in getting better in your weak subject?*

**Phase 1: Own-side argument development (sessions 1-3).** In the first session, students were introduced to the project leaders and the topic. Students were instructed to consider which side of the topic they would be inclined to argue for and then placed themselves accordingly on a described spectrum line in the classroom. Those near the middle (undecided), were engaged in conversation and eventually moved to a particular side. These sides then came to constitute the two teams the students would henceforth be members of (weak-side team and strong-side team). Teams were fairly even in size and evenly distributed by gender within sides. Once the students
had been placed on the two teams, teams moved to two separate classrooms locations with a lead
teacher present in each.

As a first activity in the new team space, students were seated in quads and asked to try to
individually describe what they believed to be the most important reason that would support their
team's position. They then wrote these reasons on post-its- one reason per post-it. Once all team
members were ready, they took turns reading the reasons to each other and, if they had reasons
that overlapped, they placed the two similar reasons together in the middle of the table. If the
reasons seemed to be overly wordy or long, teachers encouraged students to rewrite reason cards
to state a reason in the most succinct and clear way possible and to make sure each of the quad
members could recognize the reason from what was written on the card. Once all quads had
determined that all their reasons were represented, each quad shared their reasons with the other
quads, one reason per quad at a time. Assuming that all quads in the team were in agreement that
the reason was important, it was placed at the top of a team poster board - a board that would
become the gathering place for all reasons and, eventually, represent the team’s position. Teams
then reviewed and completed amendments to their poster boards. The adult leader noted how
many different reasons can underpin the same position. Students were asked to interact with the
reasons on the board by considering whether some reasons were ‘stronger’ than others and to
contemplate what made one reason stronger than another, and evidence, as able to strengthen (or
weaken) reasons, was noted as a factor by the adult leader if a student did not mention it, leading
to a discussion of its role. Students then ranked their reasons by strength, moving post-its
containing stronger reasons to the left on their board.
During the next two sessions, factual information was introduced that teams could attach to their reasons as evidence to support them. A teacher distributed lists of questions (with space under each question to write answers), introducing them as questions - the answers to which might be useful in their arguments. The list of short questions, along with the brief answers that were provided on request, are shown in Appendix 3. Students were allowed to choose from the list specific questions they would most like answers to. Students in their quads took turns choosing one question at a time and a teacher or assistant would then hand them a folder that contained a short answer to the question inside. The quad members were responsible for making sure that all members recorded answers in their own words in the space provided on their question sheets, so that these would be available for later use.

Quads then conferred to determine for each Q&A how it might serve to support (or not) a reason on their board. The idea of evidence thus was not explicitly taught, but rather arose from the discussions that the group were having about the usefulness of the answers students found in their own way, thereby earning the term ‘evidence’, rather than students being given this language by the teacher. This is an important aspect of the program - instead of being directly taught by teachers what the components of an argument are, concepts such as evidence (or later, counterargument) arise naturally as a part of discussion.

Once a group had decided which reason(s) their evidence connected to, the evidence was summarized in its simplest form on a post-it and, if agreed upon by the whole team, placed adjacent to the relevant reason on the posterboard. If necessary, students were asked to simplify the evidence text to its most important key words. If the evidence was written out in its entirety, the teacher noted, it might be difficult to identify it quickly and use it during the upcoming final
debate and an example of one question-and-answer pair, along with the student-constructed version, is this one:

**Printed Question:** What can happen when people devote all their time to one activity?

**Printed Answer:** Some people who focus on one subject might miss out on exploring others they might like. The desire to be at the “top” has also been connected to depression, anxiety, and academic dishonesty.

**Student-constructed Post-it Version:** MISS OUT AND MENTAL HEALTH RISKS

**Phase 2: Electronic dialogs (sessions 4-8).** At the beginning of session 4, the teacher leading each team asked participants how effective they felt the reasons were that had been compiled on their side of the argument. Once agreement had been reached that the reasons were strong, teachers asked the students what the next step might be to move forward toward a final debate. Students discussed amongst themselves and came to the conclusion that they needed to know what reasons the other team had and how strong they were, and the teacher supported the conclusion that they needed to find this out and that the best way to do that would be to hear from members of the other team directly. In order to do this, the teachers introduced the format of the electronic dialogs that a pair of students from one team were to conduct with a pair from the opposing team. The medium used was Google Spaces, a platform that all students at the school have access to and could be monitored as well as saved. To the extent it was possible, students were paired with a constant same-side partner for all 5 dialog sessions. The other-side pairs they conducted dialogs with rotated from one to another dialog.
In each class session, there occurred two dialogs lasting 20 minutes each. Teachers framed the ground rules around the electronic exchanges by letting students know that the google spaces would be monitored. Decorum was expected to be commensurate with in-person dialogs and students were asked to focus on ideas and not the people behind the ideas. Students were then told that the pairs would take turns with the other team in starting a dialog; pairs on one side would be designated to start and would begin the electronic debate by stating what side they were on and why. The opposing pair then discussed between themselves how to respond, and it was stressed that no response should be entered on the keyboard until the two same-side partners agreed.

In total, a given pair participated in 10 dialogs over five sessions (2 per session) with an opposing pair, with the particular pair they encountered rotating across dialogs. During these electronic dialogs, there was wait time on both sides while the opposing pair decided how to respond. Teachers reminded students that this wait time could be spent usefully by reviewing their own posterboards and question lists. They were encouraged to seek answers to every question, even if they didn’t think it would add to their position, since this information would likely be introduced by the other team and they would need to have a response to it. This meant they had two objectives in choosing a question to be answered - either to strengthen their own reason or to weaken a reason of the opposing side.

This sequence of two electronic dialogs per class and intervening wait-time (most often, used in further working to connect evidence to reasons) was repeated over five sessions. The same-side verbal exchange that occurred during the electronic dialogs enhanced the overall amount of student-to-student discourse that occurred. Students were, simultaneously,
participating, verbally and electronically, in both same-side and other-side discourse, thereby maximizing time spent in reasoned discourse.

**Phase 3: Preparing for and Conducting Whole-class final Showdown Debate (sessions 9-13).** In session 9, students met separately in their teams. They discussed what reasons they had heard in the electronic dialogs, as the basis for preparing an “Other-side” posterboard parallel to their own posterboard, presenting the other side’s position. It included any evidence they had identified that supported the other-side’s reasons.

The next step was to develop counters (counterarguments) to the reasons now displayed on the other-side board. Quads were asked to discuss and agree on the strongest counters to each of the other-side reasons and, using a new post-it color, and once whole-team agreement was reached, to place these adjacent to the reason it countered. If the counter was agreed to be a strong one that significantly weakened an opponent's reason, the post-it was to be of red color, while a weaker counter, less effective against the opponent's reason, of a pink color. The quads took turns placing counters, and the team reached agreement that duplicates had been eliminated and that the strongest counters were placed at the top.

The culminating phase of the workshop (sessions 10 & 11) consisted of a whole-class “Showdown” verbal debate, in which a participant from each team debated a member of the other team in live verbal exchange. The showdown session began with the teachers explaining the physical format as well as the rules governing the event.

The physical set-up consisted of two rows of chairs placed on each of two sides of the room, where each team assembled, with two single chairs facing each other in the center. Adjacent to each team were their prepared posterboards. In an order predetermined by the team, one member of each team assumed a position in one of the center chairs. One team opened the
debate (alternating across teams). After two minutes, hot-seat occupants were replaced by one of their teammates. Those team members seated in the front row acted as advisors to their teammate in the center. Teammates seated in the back row were allowed to pass post-it notes to the front row with suggestions, but otherwise did not participate. Half-way through the event, front and back rows reversed positions.

Once per two-minute turn, either of the two students in the center ‘hot seat’, or any member of their team, could call for a one-minute ‘huddle.’ During a huddle, the front-row debate participants and the hot-seat participant of each team would huddle together at the back of their side of the room, refer to their team’s poster-board if they wished, and discuss strategy. After one minute, the two hot-seat speakers returned to their positions and the debate resumed. The back-row participants remained seated but were allowed to pass post-it notes to their team members.

Some students were initially hesitant to assume the hot-seat position, but by the end of the two Showdown sessions, all ended up doing so. Every student therefore played the hot-seat role either once or twice during the two-session activity.

Teacher-led group reflection on their performance began the final session (session 12) and a final essay completed this (session 12). No winner was announced but all students joined in food treats and were complimented on their performance. Students were asked what they felt had been successful in their showdown performance and what had been most challenging, and teachers added to this discussion with their own comments on strengths and challenges. As a consolidating activity, in the second half of the class, students were asked to individually write a final “position piece” essay on the topic, addressed to someone unfamiliar with the topic. All students completed this task during the 25 minutes allocated to it.
Development of Coding System and Coding of Constructed Dialogs

To devise a coding system for analysis of the constructed dialogs written by the present sample, we drew on the philosophical literature noted earlier as well as our research team’s own analyses of the adult dialog that was developed and used in our previous study (Kuhn et al., 2024) and presented in Chapter 1. Some of the characteristics identified in these analyses had been noted and used in earlier research, while others were new to the present and related contemporary studies. The most fundamental one is use of Evidence, employed as a means of supporting a claim (Crowell & Kuhn, 2014); it can also be used as a means of weakening or discounting a claim. Another is Counterargument, employed for the purpose of weakening an opposing claim in a dialogic argument, although it can also be employed to identify and acknowledge a claim that figures in one’s own argument.

Our analyses, however, confirm that not all counterarguments are of equivalent force. In Table 1, we identify four types. The first, simple Disagreement, does not strictly speaking deserve the label of Counterargument at all, since it includes no argument for denying the claim, beyond asserting that it should not be accepted. What a Disagreement does do, however, is to acknowledge the opposing claim by addressing it. Failing to acknowledge or address an opponent’s claim is a common characteristic of the dialogs of novice arguers, who confine themselves to a one-sided argument that presents only their own position (Felton & Kuhn, 2001).

Genuine counterarguments that provide reasons to support disagreement can be of different types that differ in strength. A Counter-Alternative counterargument offers an alternative argument to support the conclusion being drawn, with the implication that it is a superior one, but the initial argument is never addressed, hence failing to weaken it and leaving
unfulfilled a core purpose of argument i.e. to weaken an opponent’s position (Macagno & Walton, 2015).

The two remaining counterargument types are Counter-Critique and Counter-Undermine, and both of these have the clear intent of weakening an argument. Counter-Critique does so by identifying its negative consequences. Counter-undermine is a deeper counterargument because it challenges the argument itself, as involving faulty reasoning (Macagno, Paus, & Kuhn, 2015). See Table 1 for illustrations.

The remaining categories of characteristics appearing in Table 1 go beyond simply counterargument in their intent. All were identifiable in the model dialog between adults presented in Chapter 1. The first, Question, reflects Walton’s second purpose of argumentation, to secure commitments from the opponent that can be employed to the questioner’s advantage. Occasionally a question may be simply informational, but most often in skilled argumentation it has a strategic purpose.

The next type, Concession, occurs when one of the interlocutors fulfills this objective by conceding a point. Concession is a critically important category because it signifies that the discourse has succeeded in moving one or the other participant closer in position to the other one. As such it plays a role in leading to the ultimate purpose and achievement of argumentation, that of reconciliation.

Finally, Meta statements is also a critically important category. It is defined as a statement that is about the discourse itself, in contrast to idea units that are part of that discourse. These statements were prevalent in the demonstration dialog between adult actors presented in Chapter 1. See Table 1 for an illustration from the present young adolescent student database. Although not directly advancing the argumentation, these statements do set the stage for
advancing it, by helping to establish as common ground what would do so. The example in Table 1 is a highly competent one that serves this purpose, but Meta-statements that make smaller steps in serving the purpose of establishing common ground are simpler ones observed among less skilled teen arguers such as “I never said that” or “What’s your point?”. What is notable regarding Meta-statements is the potential they have to pave the way for the previous category, Concession, which constitutes the second purpose of argumentation identified by philosophers such as Walton.

**Table 1**

*Summary and Examples of Coding Scheme for Constructed Dialogs*

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
</table>
| Counter-Alternative       | Indirectly attack the opponent’s claim by providing a different argument    | *But adolescents receive less punishment and shorter sentences in juvenile courts and don’t waste much time in jails.*
                                                                                     | (In response to: “If the government doesn’t have to open a separate juvenile court, they can save a lot of money”) |
| Counter-Critique (Direct-Counter) | Directly criticize the opponent's preceding claim as unacceptable by identifying its negative consequences | *That is incorrect because it is by law in most countries that they must be 18 to be put in adult court.*
<pre><code>                                                                                 | (In response to: “Also if they are almost an adult they should be treated like one”) |
</code></pre>
<p>| Counter-Undermine (Direct-Counter) | Direct challenge of the connection between opponent’s premises and conclusions or challenge of the premises as incorrect | <em>I see where you are coming from but since these kids were so tough to commit these crimes they should be able to understand the consequences for their actions</em> (In response to: “I disagree because kids aren’t ready for the real jail) |</p>
<table>
<thead>
<tr>
<th>Disagree</th>
<th>The weakest type of counterargument, rejecting the interlocutor’s preceding statement without justification</th>
<th>Well overall as you can see I believe that teens should not go to adult court for various reasons.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meta</td>
<td>A statement about the dialog or the act of debate itself</td>
<td>But their [there] is more evidence to support my reasoning.</td>
</tr>
<tr>
<td>Questioning</td>
<td>A statement requesting a response from the interlocutor, typically ending with a question mark</td>
<td>Could you add more to this?</td>
</tr>
<tr>
<td>Concession</td>
<td>Acknowledgement of merits of the interlocutor’s position or weaknesses of one’s own position.</td>
<td>OK I understand, they will go to real adult prison.</td>
</tr>
</tbody>
</table>

*Note. Examples are taken from the constructed dialogs in which participants are addressing the following prompt: Lex and Robyn have different positions. Lex favors adult court for teens who commit serious crimes. Robyn favors juvenile court. They are discussing the topic. They are both expert arguers and evenly matched. Write a script of what they might say.*

Two additional categories do not appear in Table 1 because they constitute second-level categorizations conducted after the initial segmentation into idea units and coding of each of those units as falling into one of the categories in Table 1. These categories are Evidence and However.
**Evidence coding.** At a second level of coding, already coded idea units were further coded as to whether they included the use of evidence to support the claim being made, in which case the unit was coded as evidence-based (EV). The evidence may have originated in documentation provided to students during the intervention (shared evidence) or be drawn from personal experience (personal evidence). In cases where students introduced evidence that was not discernably connected to an argument they were making, or the evidence was misleading or incorrect and therefore did not support their argument, the unit was coded as containing attempted but non-functional evidence.

**However coding.** 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 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. Only successful *However* pairs were included in analyses. Examples of successful and unsuccessful *However* unit-pairs appear in Table 2.

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 | Lex: They are still children
Lex: But what they have done is unacceptable and they should be punished severely* |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Unsuccessful <em>However</em> pairing</td>
<td>Adjacent statements not connected in content.</td>
<td>Lex: …if a random teenager robbed my house, I would want him to be put in a proper jail and get their proper punishment Robyn: However, they still are still kids and they need to continue their education</td>
</tr>
</tbody>
</table>

*All spelling is original to student sample*

**Coding of Constructed Dialogs for present sample.** With these categories in place, constructed dialogs that formed the data base collected from our sixth-grade sample were coded according to the coding scheme presented and illustrated in Table 1, along with the further categorization according to the two second-level categories, Evidence and *However* (Table 2). In order to mitigate bias, students were de-identified and student dialogs were randomized and numbered, with information regarding time and condition removed. The coding was done by the author and a research assistant and 30% of the dialogs were coded by both the author and the other coder, to
establish inter-rater reliability. Information regarding inter-rater reliability of coding is presented below for each outcome variable that was coded.

Each constructed dialog in the data base was first segmented into individual idea units, i.e., statements that convey a single assertion accompanied by a reason or evidence that support the claim (Felton & Kuhn, 2001). Each idea unit (after the first statement in the constructed dialog) was further coded into a category to identify the functional relation of the idea unit to the immediately preceding idea unit. An idea unit that did not qualify as an argument (i.e., did not include a reason or evidence or was off-topic completely) was defined as non-functional (NF) and was not included in the analysis. Idea units failing to convey a clear meaning or repetitions of a previous idea unit without elaboration were also excluded from the analysis. In the classification of functional/ non-functional idea units, percentage agreement between the two coders in classifying on- and off-topic idea units was 90% (Cohen’s Kappa = 0.88, p < .001).

Functional idea units were next classified as falling into one of the categories shown in Table 1. In the classification of idea units into categories, acceptable inter-rater agreement of 91% was achieved between two raters on 30% of the data, Cohen’s Kappa = 0.86 p < .001. Percentage agreement between the two coders in coding idea units as functional evidence-based units was 94% (Cohen’s Kappa = 0.89, p < .001). Only functional use of evidence was included in the analysis. All disagreements were resolved through discussion and the remaining constructed dialogs were coded by the author.

### Table 3

*An example of a constructed dialog from the present study*

<table>
<thead>
<tr>
<th>Turn</th>
<th>Hypothetical Interlocutor</th>
<th>Contribution</th>
</tr>
</thead>
</table>

32
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lex</td>
<td>I think they should be sent to regular adult prison because they commit crimes that are committed* by adults and they should be punished the same as adults for the same crimes. Also, they need to learn their lessons with real punishment not punishment which is given for stealing a chocolate bar from a convenience store.</td>
</tr>
<tr>
<td>2</td>
<td>Robyn</td>
<td>I disagree, they should go to the juvenile court because not all adults commit those crimes and some commit crimes that are even worse. Also, some people are dangerous in prison and the teens don't deserve to be with them even though they committed serious crimes.</td>
</tr>
<tr>
<td>3</td>
<td>Lex</td>
<td>You underestimate teens and their power, they are much more dangerous than you think they commit very serious crimes like robbing banks and murdering kids and adults. They deserve to be with dangerous people because they are dangerous as well.</td>
</tr>
<tr>
<td>4</td>
<td>Robyn</td>
<td>Lex, you have to remember they are just kids who had a bad education and mental support so they have gone crazy really what they need is to go to a mental health center.</td>
</tr>
<tr>
<td>5</td>
<td>Lex</td>
<td>That still doesn't cancel out all of their serious crimes. They deserved to be punished so therefore they will go to prison and then maybe after that they can go to a mental health center. Adult prisons comprise of mostly adults, so there will be no irrational behaviors.</td>
</tr>
<tr>
<td>7</td>
<td>Robyn</td>
<td>OK I understand, they will go to real adult prison.</td>
</tr>
</tbody>
</table>

*Note. Segmentation of speaker turns into individual units has been added.

All spelling is original to student sample.
Chapter 4: Results

Initial constructed-dialog performance

Our initial analysis is a descriptive one, comparing the performance of sixth-grade students in our sample to that of a demonstration of skilled argumentation conducted by skilled adult actors (see Chapter 1), for the purpose of assessing the nature and range of skill levels that individual non-expert young adolescents possess with respect to argumentation strategies and skills. In a second section of this chapter, we compare their performance over time and to that of a comparison group which did not participate in an intervention designed to maximize skill development.

In the present section, we base analysis only on the intervention group, which as we will report, turned out to be initially slightly higher performing than the non-intervention group. In addition to comparing their performance to that of the adult expert discourse sample from our earlier study, we note for purposes of comparison the performance of a contemporaneous young adolescent sample (Xiao & Kuhn, 2024), who lacked the present sample’s educational advantage and in particular their experience in argumentation and written argument as part of their academic curriculum. Doing so allows us to fulfill our objective of identifying the full range of the skill levels that individual adolescents bring to argumentation.

A total of 39 constructed dialogs from the present sample are examined, produced by members of the intervention group prior to intervention and thereby eliminating possibilities of elevated performance attributable to the intervention or simply to practice effects produced by repeated assessment. For the intervention group, the total number of idea units was 357 with a mean of 9.15 and a standard deviation of 3.38.
Presented in Table 4 is a summary of the performance displayed in this set of initial constructed dialogs with respect to each of the codes identified in Chapter 3 (seven simple codes and two second-level codes). In the first column there appears, for each of the codes, the percentage of the sample who in their initial constructed dialog ever displayed the code and in the second column the mean number of uses of the code (with standard deviations). For purposes of comparison, appearing in Table 5 are the parallel data on initial constructed dialogs reported by Xiao and Kuhn (2024) for their less advantaged sample. For this comparison group, the total number of idea units was 221, with a mean of 4.43 and a standard deviation of 1.53.

<table>
<thead>
<tr>
<th>Codes</th>
<th>% Ever-used in initial CD</th>
<th>Mean Number of Uses + SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counter-Alternative</td>
<td>85%</td>
<td>2.18 (1.59)</td>
</tr>
<tr>
<td>Counter-Critique (Direct-Counter)</td>
<td>67%</td>
<td>1.23 (1.23)</td>
</tr>
<tr>
<td>Counter-Undermine (Direct-Counter)</td>
<td>74%</td>
<td>1.85 (1.69)</td>
</tr>
<tr>
<td>Disagree</td>
<td>26%</td>
<td>0.28 (0.51)</td>
</tr>
<tr>
<td>Meta</td>
<td>21%</td>
<td>0.33 (0.70)</td>
</tr>
<tr>
<td>Questioning</td>
<td>8%</td>
<td>0.13 (0.47)</td>
</tr>
<tr>
<td>Concession</td>
<td>64%</td>
<td>1.54 (1.79)</td>
</tr>
<tr>
<td>However statements</td>
<td>69%</td>
<td>1.67 (1.71)</td>
</tr>
<tr>
<td>Evidence</td>
<td>79%</td>
<td>2.05 (1.85)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Codes</th>
<th>% Ever-used in initial CD</th>
<th>Mean Number of Uses + SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counter-Alternative</td>
<td>100%</td>
<td>3.87 (1.71)</td>
</tr>
<tr>
<td>Counter-Critique (Direct-Counter)</td>
<td>39%</td>
<td>0.43 (0.59)</td>
</tr>
<tr>
<td>Counter-Undermine (Direct-Counter)</td>
<td>17%</td>
<td>0.17 (0.39)</td>
</tr>
<tr>
<td>Disagree</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Meta</td>
<td>2%</td>
<td>0</td>
</tr>
<tr>
<td>Questioning</td>
<td>9%</td>
<td>0.09 (0.29)</td>
</tr>
<tr>
<td>Concession</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>However statements</td>
<td>9%</td>
<td>0.13 (0.34)</td>
</tr>
<tr>
<td>Evidence</td>
<td>91%</td>
<td>2.09 (1.62)</td>
</tr>
</tbody>
</table>

Table 4.  Present Study- Constructed Dialogs intervention group/ pretest  

Table 5.  Xiao & Kuhn, 2024. A-only group/ pretest
This comparison yields two broad conclusions. The present sample shows some but not all of the competencies represented in the expert adult dialog from our earlier sample (Chapter 1). Their performance is superior, however, to that of the Xiao and Kuhn (2024) less advantaged and less experienced sample, both quantitatively (number of idea units) and qualitatively (categories used).

Relative to the adult dialog, particularly notable is the minimal use of the critically important categories of Question and Meta-talk, which are used to strategic advantage in skilled argumentation to move the discourse forward. Also minimally used are the advanced counterargument types and However unit-pairs that can move the dialog forward toward integration and even reconciliation. In the Discussion chapter, we return to a fuller consideration of these differences and explanations for them. Also in the next chapter, we consider their implications, particularly with respect to the epistemological issues surrounding what adolescents understand the purposes and objectives of skilled argumentation to be and with respect to educational implications.

**Post-intervention performance**

This section presents the results of a comparison of participants’ constructed dialogs at two time points, the initial time point, examined in the preceding section of this chapter, and a second time point following the intervention described in Chapter 3. Based on past research involving this intervention, it was expected that performance would improve from first to second assessment. Also included in this analysis is a comparison of performance of these participants who took part in the intervention to a non-intervention comparison group who completed the same pre- and post- constructed dialog assessments. Independent variables included in this
analysis were thus change over time (initial and final assessments) and condition (intervention and non-intervention comparison).

A total of 39 constructed dialogs in the intervention condition and 35 in the comparison condition were completed at both pre and post assessment and are included in the analysis. The unit of analysis was an idea unit within the constructed dialog. The main dependent variables were number of idea units and the frequencies of occurrence of the categories of idea units described in Chapter 3 (Table 1). Secondary codes of Evidence use and However unit pairs were also examined.

A negative binomial analysis (Hilbe, 2011) was used as the main statistical analysis to compare performance across experimental and comparison conditions. The decision to use a negative binomial analysis was based on the distribution and nature of the data. As the count data distribution is not normal in this dataset, a standard regression analysis would not be appropriate. An equivalence of mean and variance is an assumption of the Poisson distribution but many variables in the current dataset were impacted by overdispersion. To determine which distribution would be most appropriate for each set of count variables, a one-sample Kolmogorov-Smirnov test was performed as well as examination of the mean/variance ratio. Therefore, the statistics reported are based on a negative binomial analysis, as noted, and detailed fully in Table 8.

To assess within-subject change from initial to final assessments, the nonparametric Wilcoxon signed-rank test was employed. A dependent t-test would not have been appropriate as the count data were not normally distributed. Also included in the analysis was an examination of whether a participant ever showed each of the assessed dependent variables, examined by time and condition. A McNemar test was employed in these analyses to determine significance of
change. This test was chosen as it is non-parametric and can be effectively used with count data in small samples. When the expected frequencies in cells fell below the given criteria of five, the exact binomial was used instead of the McNemar.

**Idea units** The mean number of idea units (IUs) at pretest and posttest appear by condition in Figure 1. The mean number of idea units generated by the intervention group at pretest was 9.15 (SD = 3.38) and at posttest 7.77 (SD = 2.80), a significant decline, \(z = -2.50, p = .012\). The mean number of idea units generated by the comparison group at pretest was 7.49 (SD = 2.80) and at posttest 6.77 (SD = 2.76), a nonsignificant difference. A negative binomial analysis was performed comparing conditions. After adjusting for pretest performance, there was no significant group difference in IU frequencies in posttest constructed dialogs. See Figure 1 and Appendix Table 1A.

**Figure 1**

*Mean Idea Unit (IU) frequencies in Constructed Dialogs by Condition and Time*
Counter-Alternative (CA) frequencies As the claim of the interlocutor is not addressed, Counter-Alternative is one of the weaker forms of counter argument. Although increasing over time in both groups, these increases in CA frequencies did not reach significance. The mean number of CA frequencies generated by the intervention group at pretest was 2.18 (SD = 1.59) and at posttest 2.41 (SD = 1.80), a nonsignificant difference. The mean number of idea units generated by the comparison group at pretest was 1.63 (SD = 1.57) and at posttest 1.8 (SD = 1.51), also a nonsignificant difference. Nor did the negative binomial analysis show a significant group difference after adjusting for pretest performance. See Figure 2 and Appendix able 2A. In the comparison group at pretest, the percentage of students who ever used CA was 74.3% which increased to 88.6% at posttest. In the intervention group, the percentages increased from 84.6% at pretest to 87.2% at posttest. While the number of times a counter-alternative was ever used by participants within both conditions increased, the McNemar test was not statistically significant within either conditions.

Figure 2

Mean Counter-Alternative (CA) frequencies in Constructed Dialogs by Condition and Time
**Counter-Critique (CC) frequencies**  The mean number of CC frequencies generated by the intervention group at pretest was 1.23 (SD = 1.22) and at posttest 0.63 (SD = 0.85), a significant decline, ($z = -2.808$, $p = .005$) The mean CC frequencies generated by the comparison group at pretest was 1.23 (SD =1.06) and at posttest 0.46 (SD = 0.70), also a significant decline, ($z = -3.55$, $p <.001$). A negative binomial analysis was performed; these results indicated that while there were no group significant group differences, the covariate (pretest) had a significant effect on the frequency outcome at posttest ($\hat{\beta} = 0.329$, $p = .04$). See Figure 3 and Appendix table 3A. The McNemar test conducted to assess the number of times a counter-critique (CC) was ever used by a student within conditions revealed a statistically significant decrease in ever-used CC’s within both the comparison ($p = .001$) and intervention condition ($p = .21$) the opposite of the hypothesized direction.

**Figure 3**
Counter-Critique (CC) frequencies in Constructed Dialogs by Condition and Time

Counter-Undermine (CU) frequencies The mean number of CU frequencies generated by the intervention group at pretest was 1.85 (SD = 1.69) and at posttest 1.69 (SD = 1.24), a nonsignificant difference. The mean number of idea units generated by the comparison group at pretest was 1.29 (SD = 1.45) and at posttest 1.17 (SD = 1.27), a nonsignificant difference. A negative binomial analysis was performed; after adjusting for pretest performance, there was no significant group difference in the frequencies of CU in the posttest constructed dialogs. See Figure 4 and Appendix table 4A. In the comparison group at pretest, the percentage of students who ever used CU was 60% which increased to 69% at posttest. In the intervention group, the percentages increased from 74% at pretest to 85% at posttest. While the number of times a
counter-undermine was ever used by participants within both conditions increased, the McNemar test was not statistically significant within either conditions.

**Figure 4**

*Mean Counter-Undermine (CU) frequencies in Constructed Dialogs by Condition and Time*

**Evidence frequencies**

The mean number of Evidence units generated by the intervention group at pretest was 2.05 (SD = 1.85) and at posttest 2.77 (SD = 2.37), a significant difference ($z = 2.07$, $p = .039$) with a small-medium effect size, calculated using Cohen’s $d$ ($d = 0.33$). The mean number of Evidence units generated by the comparison group at pretest was 0.94 (SD = 1.39) and at posttest
1.51 (SD = 2.00), a nonsignificant difference. Evidence usage increased over time in both conditions with participants in the experimental condition demonstrating a higher mean frequency of use of Evidence in the posttest constructed dialogs. A negative binomial analysis was performed; these results indicated that while there were no significant group differences, the covariate (pretest) had a significant effect on the frequency outcome at posttest ($\hat{\beta} = 0.295$, $p = .001$). See Figure 5 and Appendix table 5A. Despite the increase in ever-used Evidence from 49% at pretest to 60% at posttest in the comparison group, the McNemar test was not statistically significant within either condition. It is noteworthy, however, that 79% of students in the intervention both at pretest and posttest ever used evidence in their writing. I will consider this further in the discussion chapter.

**Figure 5**

*Mean Evidence (EV) frequencies in Constructed Dialogs by Condition and Time*

![Graph showing mean evidence frequencies in constructed dialogs by condition and time.](image)

**However frequencies** As seen in Figure 6, the mean number of *However* unit pairs generated by the intervention group at pretest was 1.66 (SD = 1.71) and at posttest 0.84 (SD = 0.81), a
significant decline, \((z = -2.37, p < .018)\). The mean number of *However* statement units generated by the comparison group at pretest was 1.03 (SD = 1.01) and at posttest 0.57 (SD = 1.00), also a significant decline, \((z = -2.06, p = .040)\). A negative binomial analysis was performed; after adjusting for pretest performance, there was no significant group difference in the frequencies of *However* statements in the posttest constructed dialogs. In the comparison group at pretest, the percentage of students who ever used *However* statements was 60\% which decreased to 37.1\% at posttest (see Appendix Table 6A). In the intervention group, the percentages decreased from 69.2\% at pretest to 61.5\% at posttest. While the number of times a *However* statement was ever used by participants within both conditions decreased, the McNemar test was not statistically significant within either condition.

**Figure 6**

*Mean *However* Statement (HS)* frequencies in Constructed Dialogs by Condition and Time*
Usages were very low for the remaining four codes – Disagree, Meta, Questioning and Concession. Means, standard deviations and ever-used percentages are summarized in Table 6-1 for the intervention group and in Table 7A for the comparison group (see Appendix). Summarized for each of the categories that were analyzed across time are percentages ever showing the category at the two times (and whether the change reached significance) and the mean use at each time.

The mean number of ‘Disagree’ units generated by the intervention group at pretest was 0.28 (SD = 0.51) and at posttest 0.10 (SD = 0.31), a nonsignificant difference. The mean number of ‘Disagree’ units generated by the comparison group at pretest was 0.09 (SD = 0.24) and at posttest 0.11 (SD = 0.32), a nonsignificant difference. In the comparison group at pretest, the percentage of students who ever used Disagree was 9% which increased to 11% at posttest. In the intervention group, the percentages decreased from 26% at pretest to 10% at posttest.

The mean number of ‘Meta’ units generated by the intervention group at pretest was 0.33 (SD = 0.70) and at posttest 0.28 (SD = 0.56), a nonsignificant difference. The mean number of ‘Meta’ units generated by the comparison group at pretest was 0.46 (SD = 0.92) and at posttest 0.37 (SD = 0.77), a nonsignificant difference. In the comparison group at pretest, the percentage of students who ever used ‘Meta’ was 23% which remained at 23% at posttest. In the intervention group, the percentages increased from 21% at pretest to 23% at posttest.

The mean number of ‘Questioning’ units generated by the intervention group at pretest was 0.13 (SD = 0.47) and at posttest 0.08 (SD = 0.35), a nonsignificant difference. The mean number of ‘Questioning’ units generated by the comparison group at pretest was 0.09 (SD = 0.28) and at posttest 0.14 (SD = 0.36), a nonsignificant difference. In the comparison group at pretest, the percentage of students who ever used Questioning was 9% which increased to 14% at
posttest. In the intervention group, the percentages decreased from 8% at pretest to 5% at posttest.

The mean number of ‘Concession’ units generated by the intervention group at pretest was 1.54 (SD = 1.79) and at posttest 0.74 (SD = 0.95), a nonsignificant difference. The mean number of ‘Concession’ units generated by the comparison group at pretest was 1.06 (SD = 1.16) and at posttest 0.89 (SD = 0.90), a nonsignificant difference. In the comparison group at pretest, the percentage of students who ever used Concession was 57% which increased to 60% at posttest. In the intervention group, the percentages decreased from 64% at pretest to 54% at posttest.

Again, for purposes of comparison, appear in Table 6-2 parallel data reported by Xiao and Kuhn (2024) for their sample. Notable in this comparison is the fact that their post-intervention performance reaches roughly the same levels as the present sample’s pre-intervention performance, while the present sample in contrast does not show improvement as a result of the intervention. Discussed in the next chapter is whether this difference reflects an asymptote in achievement levels among the present sample - a conclusion, if warranted, that has important educational implications given the present group’s continuing minimal use of the important categories identified in Chapter 1 at the expert level and critical to skilled argumentation.

**Table 6-1**

*Intervention Group Summary (Present study)*

<table>
<thead>
<tr>
<th>Codes</th>
<th>Intervention Group /Pre-intervention ever-used</th>
<th>Mean &amp; SD</th>
<th>Intervention Group /Post-intervention ever-used</th>
<th>Mean</th>
<th>McNemar</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Codes</td>
<td>Pre-intervention ever-used</td>
<td>Mean &amp; SD</td>
<td>Post-intervention ever-used</td>
<td>Mean &amp; SD</td>
<td>McNemar</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-----------------------------</td>
<td>-----------</td>
<td>-----------------------------</td>
<td>-----------</td>
<td>---------</td>
</tr>
<tr>
<td>Counter-Alternative</td>
<td>100%</td>
<td>3.87 (SD 1.71)</td>
<td>96%</td>
<td>2.30 (SD 1.40)</td>
<td>NS</td>
</tr>
</tbody>
</table>

**Table 6-2**

*Xiao & Kuhn (2024) Summary*
As seen in Tables 6.1 and 6.2 above, the percentages of students in the Xiao and Kuhn sample who ever used Counter-Undermine, Concession, and However Statements remained lower at both pre-test and post-test than in the present study. In addition, the mean number of idea units in the Xiao and Kuhn sample were lower at pre-test (4.48) and at post-test (4.87) than in the present study- at pre-test (mean 9.15) and close to the mean in the present study mean at post-test (4.43).

In both samples, Questioning and Meta statements remained low across time while Counter-Alternative remained high across time. The implications of these findings will be discussed in the next chapter. See Table 7 for statistical summary.

**Table 7**

*Estimation Results of Negative Binomial/ Poisson Regressions on Post-intervention frequencies*
<table>
<thead>
<tr>
<th>Response Variable</th>
<th>Estimates</th>
<th>Std Error</th>
<th>Conf. Interval</th>
<th>IRR</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Idea Units</strong></td>
<td>Intercept 1.68</td>
<td>0.41</td>
<td>(0.87 to 2.48)</td>
<td>0.93</td>
<td>.78</td>
</tr>
<tr>
<td></td>
<td>Condition -0.071</td>
<td>0.26</td>
<td>(-0.575 to 0.434)</td>
<td>0.87</td>
<td>.64</td>
</tr>
<tr>
<td></td>
<td>Pretest 0.04</td>
<td>0.04</td>
<td>(0.04 to -0.04)</td>
<td>1.19</td>
<td>.05</td>
</tr>
<tr>
<td><strong>Counter-Alternative</strong></td>
<td>Intercept 0.44</td>
<td>0.29</td>
<td>(-0.14 to 1.01)</td>
<td>0.87</td>
<td>.64</td>
</tr>
<tr>
<td></td>
<td>Condition -0.14</td>
<td>0.30</td>
<td>(-0.722 to 0.442)</td>
<td>1.19</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>Pretest 0.17</td>
<td>0.09</td>
<td>(0.04 to -0.04)</td>
<td>1.19</td>
<td>.05</td>
</tr>
<tr>
<td><strong>Counter-Critique</strong></td>
<td>Intercept -0.97</td>
<td>0.36</td>
<td>(-1.67 to -0.27)</td>
<td>0.74</td>
<td>.47</td>
</tr>
<tr>
<td></td>
<td>Condition -0.29</td>
<td>0.41</td>
<td>(-1.09 to 0.50)</td>
<td>0.74</td>
<td>.47</td>
</tr>
<tr>
<td></td>
<td>Pretest 0.33</td>
<td>0.16</td>
<td>(0.01 to 0.64)</td>
<td>1.39</td>
<td>.04</td>
</tr>
<tr>
<td><strong>Counter-Undermine</strong></td>
<td>Intercept 0.34</td>
<td>0.28</td>
<td>(-0.20 to 0.88)</td>
<td>0.74</td>
<td>.35</td>
</tr>
<tr>
<td></td>
<td>Condition -0.29</td>
<td>0.32</td>
<td>(-0.91 to 0.33)</td>
<td>0.74</td>
<td>.35</td>
</tr>
<tr>
<td></td>
<td>Pretest 0.09</td>
<td>0.09</td>
<td>(-0.09 to 0.27)</td>
<td>1.09</td>
<td>.33</td>
</tr>
<tr>
<td><strong>Evidence</strong></td>
<td>Intercept 0.29</td>
<td>0.28</td>
<td>(-0.25 to 0.83)</td>
<td>0.75</td>
<td>.35</td>
</tr>
<tr>
<td></td>
<td>Condition -0.28</td>
<td>0.31</td>
<td>(-0.89 to 0.32)</td>
<td>0.75</td>
<td>.35</td>
</tr>
<tr>
<td></td>
<td>Pretest 0.26</td>
<td>0.90</td>
<td>(0.12 to 0.47)</td>
<td>1.34</td>
<td>.001</td>
</tr>
<tr>
<td><strong>However Statements</strong></td>
<td>Intercept -0.30</td>
<td>.32</td>
<td>(-0.92 to 0.32)</td>
<td>1.08</td>
<td>.52</td>
</tr>
<tr>
<td></td>
<td>Condition -0.35</td>
<td>0.37</td>
<td>(-1.08 to 0.38)</td>
<td>0.71</td>
<td>.35</td>
</tr>
<tr>
<td></td>
<td>Pretest 0.08</td>
<td>0.12</td>
<td>(-0.16 to 0.32)</td>
<td>0.71</td>
<td>.35</td>
</tr>
</tbody>
</table>
Chapter 5: Discussion

Purpose of the dissertation

The purpose of this dissertation was, building upon two previous studies (Kuhn et al., 2024; Xiao & Kuhn, in press), to assess the efficacy of constructed dialog in an educationally advantaged environment as a tool for assessing individual dialogic argument skill both prior to and following an intervention known to support argument skill development. More specifically, the question was posed as to whether the intervention in the present setting might reveal more advanced skill levels that would make skill-level distinctions clearer, thus enriching evaluation of the potential of the constructed dialogs as an assessment tool in measuring students’ skill in argumentation in both research and applied settings. Lastly, the extent to which results from the performance on the constructed dialog task may be related to participants’ levels of epistemological understanding of argumentation is also examined.

Summary of results

Considering the variables highlighted in Chapter 1 as being indicative of advanced argumentation abilities (see Table 1) and illustrated in the expert actor dialogs (see Appendix 1), participants in the present study demonstrated considerable competence in several higher-order argumentation skills, particularly in the use of Evidence. Nonetheless, the frequency of usage of other key argumentation skills, notably Meta and Questioning, remained low both at pre-test and post-test - even the students in the present study who demonstrated high frequency scores on several variables were still not at asymptote across all skills, as evidenced by the low frequency scores for the dimensions noted above. Thus, critical skills, in particular metacognitive abilities, remained minimal and, in some cases, completely absent. To examine the impact of this
discrepancy on argumentation education, more development of the constructed dialog assessment, as well as additional research, will be necessary, as discussed further in this chapter.

In continuing work, our team is comparing constructed-dialog individual performance to other indicators of argumentive skill. One is the traditional measure of an individual argumentive essay (Fraguada, Bruun, & Kuhn, in preparation). The other is comparison of constructed dialog individual performance to actual dialogic performance conducted between opponents (Bruun, Xiao, & Kuhn, in preparation). Because this comparison is between the performance of an individual and performance of a pair, statistical comparison is not appropriate. The comparison is nonetheless illuminating. In dialogs conducted between opposing pairs (like those described in the present intervention), compared to the same participants’ individual constructed dialogs, number of idea units in the dialogs were almost identical, at about five. The actual dialog, however, was created by four individuals (two same-side pairs holding opposing positions), while the constructed dialog was created by just a single individual. Furthermore, comparison of the classifications of these idea units into the categories in Table 1, showed frequencies that a same-side pair in the actual dialog produced to be only about half the frequency the individual produced in their constructed dialog. Put differently, then, the individual in a constructed dialog produced frequencies of use of even the most demanding categories, such as counter-undermine, equal to those produced by the two pairs (four individuals) participating in the actual dialog.

The point of this comparison is to support the constructed dialog assessment as yielding a high-end assessment of an individual’s competence in argumentation – a finding having both theoretical and practical implications. Before proceeding to these, we consider the intervention findings.
Limitations in interpreting change across time

Data indicated minimal improvements from pre-test to post-test in the intervention group, contrary to expectation and previous research. There exist several potential interpretations. One is that the intervention itself was too limited (in hours). A further possibility is that it was not concentrated in time, contrary to previous successful interventions, instead meeting only once or twice a week. The previously noted Xiao and Kuhn (2024) intervention with students of a similar age, for example, was conducted over an intensive two-week period, with 4 hours daily. The combination of these factors may have prevented students in the present study from fully integrating, and appreciating, the skills highlighted in the intervention.

In addition, the intervention was a program that stood apart from the daily (and graded) school curriculum; therefore, students may have felt less motivated to truly engage with the material and seen less connection between their educational program and the intervention program. Lastly, there was a slight pre-test difference between participants in the intervention and comparison groups. This difference likely reflects that students who were already somewhat stronger in the relevant skills were more likely to enroll than weaker students who may have benefitted most from the experience. Students who were viewed by their families as having the capacity and discipline to handle an additional programming element were the ones more likely to enroll. The families who chose to enroll their children into the intervention were clearly motivated to have their children engage with the intervention program while students for whom the intervention may have been most beneficial never enrolled. While there was no maximum enrollment number, we did not reach out to the students who had not enrolled and therefore encountered this selection bias. This selection bias impacted the overall study and makes generalizations based on results difficult.
In addition, placing the pre-intervention constructed dialogs at the beginning of the school year may not have been the ideal time, especially as students transition from 5th to 6th grade which, at this school, meant entering a new campus, encountering new teachers and receiving letter grades for the first time. In addition, all the students took the pretest in the first week of school. Not only were the students palpably nervous as they had all transitioned to a new middle school campus but they had limited experience with assessments and felt the stakes of any assessment they encountered were extremely high. While the pretest wasn’t graded, this wasn’t explicitly stated to the students, and they were given both pre-test and post-test assessments in their regular classroom settings. Therefore, the expected rigors of 6th grade impacted the effort the students put into the pretest, evidenced by the number of idea units produced by both groups, which fell slightly over time. Over 16 weeks later, students had grown comfortable, understood the parameters of assessments and had less motivation to produce their best work. While the impact of increased motivation in the pretest environment was evident across both conditions, the selection bias was apparent here as the comparison group mean remained lower at pretest than the intervention group at posttest. A final possibility is that participants from this sample had already reached an asymptote in argumentation skill level. This interpretation, however, seems unlikely given their relatively infrequent usage of the more advanced skills, and in any case cannot be seriously entertained without ruling out other interpretation.

**Idea Units.** An indicator of ceiling effect in the idea units is that the number of idea units produced by both groups in the present study were higher at pretest than those produced by students of a similar age at posttest in Xiao and Kuhn’s, (2024) study. The number of idea units produced by the students in the present study were closer to those produced by adults in Zavala’s
This suggests that a ceiling effect could have impacted the present study. Given the juxtaposition of score declines in some variables but improvements in others, it is possible that a ceiling effect was caused by the time constraints of the intervention rather than an indication of the capacity of the students involved. Again, however, no definitive conclusions can be drawn.

**Evidence.** Evidence usage increased over time in both conditions with participants in the experimental condition demonstrating a statistically significant increased mean frequency of use of Evidence in the posttest constructed dialogs. The focus on the use of evidence in the intervention was explicit, and results suggest that the naming of this skill during instruction as well as the focus on the gathering of evidence to support claims over the intervention process led to gains observed in the intervention group—a finding that is consistent with prior research (Kuhn & Moore, 2015; Macagno, 2016). Given the central role that evidence plays in argumentation, the support that these results lend is of note.

**Meta-statements.** The scores for “Meta-statements” were higher at pretest in the present study than the level attained at posttest in the Xiao and Kuhn study. This difference indicates that students in the present study implicitly understood the value of dialogic at a deeper level. This use of meta-statements was not prompted by the intervention as the difference in results were already present at the pretest. Rather, this presents an interesting window into the epistemological awareness of the students at a rigorous educational institution and indicates that students in the present study possessed a stronger epistemological understanding of what it means to argue than did their counterparts of a similar age. While these statements did not always directly advance the arguments, they did support the communicative experience needed for opponents to find common ground. While it might be that the students in the Chinese
comparison group might have been less exposed to dissent and discourse, the growth within that sample from pre-test to post-test was also limited.

**Questioning.** The frequencies for the ‘Questioning’ code were low across studies and conditions, both pre and post intervention. As previous research has indicated, questioning within collaborative peer-groups supports the activation of prior knowledge, helps learners co-construct knowledge (Chin & Osborne, 2010) and resolves areas of confusion. While questioning might, in some cases, mainly serve the purpose of clarifying or informing, in the constructed dialogs there is an additional layer- the act of questioning is crucial when self-generated for the purpose of gaining knowledge and particularly other perspectives (Scardemalia et al., 1992) The results from this study indicate that while metacognition and epistemological awareness were emerging skills in this sample, students, across time and conditions, for the most part lacked the ability to take the extra strategic step of questioning in order to advance argumentation goals. This result raises the question of how these skills, within a dialogic context, might develop naturally and what aspects of a program could be targeted to grow the skill of questioning.

**Theoretical Implications**

Theoretically, the present study contributes to the literature on the development of argument skill, and to an equal extent to the now extensive literature on development of epistemological understanding, specifically with respect to the purposes and objectives of argumentation. When an individual constructs a piece of argumentative discourse, they reveal what they believe to be essential characteristics that such discourse possesses – knowledge of considerable potential value to theorists and educators alike. Argumentation skill is now prized as an educational outcome, as well as serving as a vehicle for achieving positive educational outcomes more broadly. This makes it doubly important to identify as fully as possible the nature of this epistemological understanding and to nurture it. This is especially so given its impact on practice. People are disposed to engage in the
demanding intellectual practice of skilled argument only to the extent that they appreciate its value (Kuhn, 1991).

Without fully grasping the value of and need for argumentation, engagement is likely to remain low, meaning that little opportunity for practice and growth occurs. Younger students who exhibit less effective argument strategies or none at all are likely to lack the motivation as well as the epistemological awareness and understanding of the argumentation process needed to support its practice. The present findings on the one hand show notable initial competence among young adolescents, while on the other hand make apparent the need for further growth, even among the educationally privileged population we worked with. Further research is needed to generalize the present findings to a broader range of age groups and populations. In addition, more research is needed to ascertain to what extent the kind of discourse-based intervention employed in the present study can support skill development in optimal ways among different populations.

**Educational Implications for Assessment and Intervention**

In addition to its theoretical contribution, the present study lends important support to the use of constructed dialogs as a novel assessment tool in measuring individual argumentation skill. Further research is needed to establish the extent to which a constructed dialog can serve as a sensitive measure of an individual’s skill in argumentive discourse. This is especially true in relation to other more traditional assessments of argumentation skill, most notably the individual essays that we are now studying and comparing to constructed dialogs as assessment tools. Our further ongoing work in examining and comparing differences between traditional essays and constructed dialogs should reveal the extent to which constructed dialogs are superior in revealing the range and discriminability of individuals’ skill-level distinctions.
Another direction for further investigation is the employment of constructed dialogs not only as an assessment tool but as a teaching tool. Would repeated practice in constructing such dialogs help to support the rapid perspective-taking shifts that skilled argumentation requires if it is to be effective in addressing and understanding another person’s view? The creator of a constructed dialog is creating and coordinating two perspective single-handedly.

One of the goals of the discourse-based curriculum used in the present study is to introduce students to collaborative argumentation and, thereby, support the development of critical thinking skills within a dialogic framework. In addition, the program aims to help students see the value of reasoned argumentation and may motivate them to apply these skills to other topics and issues. As part of this experience, constructed dialogs present an added opportunity for the creator to gain insights into how opposing perspectives differ and ultimately how they might become integrated. In addition, inherent to the structure of the constructed dialog task is the need to take multiple positions - not always a condition necessary in traditional assessments such as essays. Educators may consider the benefits of this type of task in which students are forced to consider other perspectives- even if it means opening up the traditional classroom assessments to more collaborative and engaging methods. This enables thinking to become active and social in nature with an exchange of ideas present, even when those ideas originate within the same person. A discourse-based approach may well be honed in other ways to support students at varying entry points of ability. In the present study’s advantaged sample, initial competence didn’t preclude development – an important implication both for teachers who work in high-rigor institutions and well as for researchers who are hoping to refine the approach to fit a variety of settings.
The language around argumentation is rapidly shifting in the current political climate and children are particularly impacted by the rhetoric of the adults around them as discourse becomes more polarizing and, at times, even aggressive. Knowing what they think and how others may respond in, and to, the process of argumentation is therefore increasingly crucial to make children ready to be engaged citizens. Consideration should be given to the impact of these trends on school-aged children, highlighting the importance of argumentation skills becoming a core part of a middle-school curriculum, whether integrated as a part of instruction, as in ‘Argue With Me’ or as an assessment of individual skill in argumentation.


Bruun, S., Xiao, S., & Kuhn, D. *Assessing individual competence in discourse* [Manuscript in preparation].


Ennis, R. H. (2015). *The palgrave handbook of critical thinking in higher education.* [https://doi.org/10.1057/9781137378057.0005](https://doi.org/10.1057/9781137378057.0005)


https://doi.org/10.1080/07370008.2013.830618


http://doi.org/10.1080/09500780802152499


https://doi.org/10.1007/s11245-014-9271-6


https://doi.org/10.1007/s11145-020-10117-2


65

Roderer, T., & Roebers, C. M. (2014). Can you see me thinking (about my answers)? using eye-tracking to illuminate developmental differences in monitoring and control skills and their relation to performance. Metacognition and Learning, 9(1), 1-23.  
https://doi.org/10.1007/s11409-013-9109-4


https://doi.org/10.1016/j.cedpsych.2019.101809


1a. Dialogic video text (B=Ben/ S=Sam)

B: I believe you have to send illegal immigrants back. They have no right to be in the US.
S: Why do you say that?
B: Because they broke the law. The US is a country based on laws and rules.
S: But these laws were made by the people of this country. The laws could be changed.
B: But that’s not what we’re talking about here. We’re talking about what should happen under the existing laws.
S: Okay, I’ll accept that. But then what about those who entered as children? They did not knowingly break the law.
B: Okay then let us limit our discussion to those who knowingly came as adults- they must go back.
S: Will that accomplish anything? They’ll just try to come back.
B: Not if there is a secure border. It’s supposed to being built right now. That will solve the problem of people bringing their children illegally and the problem of children growing up illegally in the US can go away.
S: But is a border wall realistic? How much would that cost?
B: The estimates say a totally secure wall would be very difficult and costly to build. So I agree it probably won’t happen at least anytime soon.
S: So wouldn’t that money on border walls and border guards be better spent on all the
needs of people in the USA?

B: Maybe, but then we haven’t solved our problem of what to do about people entering illegally.

S: But I believe that many of them need to come because they are escaping danger in their own countries.

B: Then they can apply as refugees. There’s a program for that.

S: But maybe they are in immediate danger.

B: But maybe they’re not and only pretending.

S: Refugees are subject to long investigations. And anyway, if they are applying legally as refugees they aren’t part of the problem of illegal entering that we’re talking about—people who aren’t necessarily in danger and enter illegally just because they want a better life.

B: They should be refused entry.

S: But what if they are already here? Why send them back?

B: One reason is that they are taking away the jobs from Americans.

S: What is the evidence for this?

B: Unemployment rates are high. Lots of Americans need those jobs.

S: But only in some fields. Immigrants are willing to take jobs Americans don’t want, just to get a start.

B: But what if they don’t get a job? They will likely become a burden to the US. The US may even have to give them job training and educate their children, which is very expensive. They will have to learn English.

S: But then they can contribute and become useful members of US society.
B: Does that happen often?

S: Yes, there are many examples of immigrants who contribute.

B: But why should they come? Why not stay in their own country and contribute to that country?

S: Because that’s not the American way. America is a country of immigrants. Do you know what is written on the Statue of Liberty to welcome all who enter? “‘Give me your tired, your poor…”

B: But if the US let’s anyone in who wants to come, it will be too hard to give them all the help they need to get started

S: But as I said, a great many succeed. Why not give anyone who wants it a chance to come and work hard to succeed here?

B: If it’s too easy to get into the US, more and more people will be encouraged to come. It could easily get out of control and people born in the US or already here wouldn’t have their needs met.

S: So a system needs to be set up to decide who gets priority to come to the US.

B: But the US does have a system for deciding the cases of people who apply to enter legally. It may not be the best or fairest one and isn’t working very well at the moment.

S: You have a point. So maybe we should put all efforts into figuring out a better system for people who want to come and help them to do it legally.

B: I agree, that would be best. So maybe our disagreement isn’t so great. But there are many factors to consider – we agree on that – so the task isn’t going to be an easy one.
1b. Monologic video text:

Sam: I believe people who enter the US illegally should be allowed to stay here.

Immigration laws and rules were made by the people of this country. If Americans made their own laws, that means they can also be changed. If the actual law itself is the issue, then we can have that amended so it no longer becomes about people breaking the law or doing something illegal by simply crossing a border. And sending home those who enter illegally won’t accomplish anything. They’ll just come back. If they aren’t able to cross the border on the first try, they will likely try again until they are successful. And building a totally secure wall would cost a huge amount, if it even could be done. That money that would go towards a wall or stronger border, could do a lot for Americans. So wouldn’t that money on border walls and border guards be better spent on all the many needs of people in the US?

Also, many of those who do try to come do so because they are escaping danger in their countries. They want to come to the US to find safety. They are subject to investigations to ensure their stories are true. Immigrants help the US because they are willing to take jobs Americans don’t want. A lot of positions are filled by immigrants because Americans wouldn’t take them. Having immigrants work in the US is good. They often contribute to and become useful members of US society. There are many examples of immigrants who become doctors, lawyers, scientists. America is a country of immigrants- it’s written on the Statue of Liberty to welcome people: “Give me your tired, your poor…” Why not give anyone who wants it a chance to come and work to
succeed? The US has a system for deciding the cases of people who apply to enter legally, even if it isn’t perfect and isn’t working very well at the moment. There are times when immigrants are welcome into the country, and when they are questioned. So let’s put effort into figuring out a better system for people who want to come and help them to do it legally.

Ben: I believe people who enter the US illegally should be sent back. They have no right to be in the US. They broke the law, so long as we assume they entered as adults and knew they were breaking the law. The US is a country based on laws and rules. A secure border will solve the problem. It’s supposed to being built right now. That will solve the problem of people bringing their children illegally; and the problem of children growing up illegally in the US can go away. But I know a totally secure wall probably won’t be achieved soon. It would be very difficult and costly to build. But that is logistics, in the end this issue remains, and it needs to be acknowledged. We still have to solve the problem of what to do about people entering the US illegally. Some say they are fleeing from danger, but maybe they’re not and only pretending. There is no way of knowing- but what we do know is that crossing the border the way they do is illegal. If they are coming just because they want a better life, they should be refused entry. Breaking the law because you are unhappy/ want better circumstances is not an acceptable reason to break the law. They are taking away jobs from Americans who need them. And if they don’t get a job, they will become a burden to the US. The US may have to give them job training and educate their children, which is very expensive.
They have to learn English? Why don’t they just stay in their own country and contribute there? There is only harm that comes from letting people cross the border illegally.

If the US lets anyone in who wants to come, it could be too much for the US to give them all the help they need to get started. More and more people will be encouraged to come. It could get out of control and people born in the US or already here wouldn’t have their needs met. So we can’t let everyone in and we need to figure out the best system for determining who is allowed to come and how they should do it legally.
2. Constructed Dialog task prompt:

**Important definitions:**

*Juvenile:* Relating to young people, usually under 18 years of age.

*Juvenile Court:* a court of law responsible for the trial of children under a specified age (18 in most countries).

**Instructions:**

In the Google Form, complete the following prompt:

*Write a Dialog (conversation between 2 people)*

1) **Read this before you begin writing:**

*Lex and Robyn have different positions. Lex favors adult court for teens who commit serious crimes. Robyn favors juvenile court. They are discussing the topic. They are both expert arguers and evenly matched. Write a script of what they might say.*

2) **Begin your script in the Google Form like this:**

LEX: I think they should be sent to regular adult prison because…(complete)

ROBYN: I disagree because…

Continue in the same way, filling in what each might say.

LEX

ROBYN:

LEX:

ROBYN:

etc.

You have 15 minutes to complete this task.
3. Questions & Answer Sheet:

Q: You are very good in one school subject and don’t do well in another. Should you put most
time and effort in being at the top of the class in your strong subject or in getting better in your
weak subject?

**Questions about most effort in strong subject**

1. Do children who devote much time to a particular skill sometimes achieve extraordinary
skill in it?
   
   Yes, chess, tennis, and music are all examples.

2. How long does it take to become an expert in a subject?

   *Research shows that it takes about 10,000 hours, or approximately 10 years of intense practice,
to become an expert in a subject (for example, math).*

3. Does high achievement in school raise a student’s self-esteem?

   *Research shows that doing very well in school may increase the self-esteem of a student, making them feel more positive about themselves.*

4. Are students who excel in one subject more likely to get scholarships than students who
are good, but not excellent, in all subjects?

   *There are many scholarships for students who are exceptional in a specific academic area, such as the Caroline Bradley Scholarship for students who perform very highly in at least one subject. 30 middle school students receive this scholarship.*

5. Do teens' life goals change over time?

   *Yes, a study found that two thirds of young teens changed their minds about interests and goals by their late teens.*
6. Do different students have different types of intelligences?
According to Psychologist Howard Gardner, there are 9 different types of intelligences and different people have different combinations of intelligences. For example, one person can be very strong in logical-mathematical intelligence but weak in verbal-linguistic intelligence.

7. What happens when students set their goals based on comparing themselves to others?
Students who set goals by comparing themselves to others may avoid taking risks so as not to risk losing their standing.

8. What can happen when people devote all their time to one activity?
Some people who focus on one subject might miss out on exploring others they might like. The desire to be at the “top” has also been connected to depression, anxiety, and academic dishonesty.

Questions about most effort in weak subject

9. Do people try harder when they don’t do well at something?
Research shows that because harder tasks require more effort, people will try harder up to a certain point to complete a hard task.

10. Is perseverance related to success?
Research has suggested that perseverance, or continuing to try despite failures, may be the most important factor in success.

11. How do colleges judge grades in admitting students?
According to the National Association for College Admission Counseling (NACAC), having good grades in all courses is among the most important factors colleges consider.

12. What is a growth mindset and how can it affect learning?
People with a growth mindset believe they can improve their performance by putting in effort. They welcome challenges because they see failure as an opportunity to grow and learn from experience.

13. What happens when students set goals for their learning?
When students set goals based on what they will learn, they are more likely to seek challenges and see failure as a step toward mastery.

14. Does failure always motivate people to succeed?
Failure does not always motivate people to succeed. If a person believes ability is fixed, failure may cause them to put forth less effort and even give up.

15. Does perseverance guarantee success?
Perseverance may be very important to success, but psychologists say it is only part of the equation. A person must also have passion for what they do to be good at it.

16. How does failure affect self-esteem?
One study found that students experiencing failure scored lower on self-esteem tests.
Table 1A

Negative Binomial (Constructed Dialogs) Idea Units using pretest as covariate

Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>B</th>
<th>Std. Error</th>
<th>Lower</th>
<th>Upper</th>
<th>Wald Chi-Square</th>
<th>df</th>
<th>Sig.</th>
<th>Exp(B)</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.678</td>
<td>.4113</td>
<td>.872</td>
<td>2.484</td>
<td>16.655</td>
<td>1</td>
<td>&lt;.001</td>
<td>5.357</td>
<td>2.392</td>
<td>11.995</td>
</tr>
<tr>
<td>[Group=1.00]</td>
<td>-0.071</td>
<td>.2574</td>
<td>-.575</td>
<td>.434</td>
<td>.075</td>
<td>1</td>
<td>.784</td>
<td>.932</td>
<td>.563</td>
<td>1.543</td>
</tr>
<tr>
<td>[Group=2.00]</td>
<td>0.0</td>
<td>.2574</td>
<td>.434</td>
<td>.075</td>
<td>-.575</td>
<td>1</td>
<td>.784</td>
<td>.932</td>
<td>.563</td>
<td>1.543</td>
</tr>
<tr>
<td>IU_pre</td>
<td>0.040</td>
<td>.0408</td>
<td>-.040</td>
<td>.120</td>
<td>.952</td>
<td>1</td>
<td>.329</td>
<td>1.041</td>
<td>.961</td>
<td>1.127</td>
</tr>
<tr>
<td>(Scale)</td>
<td>1b</td>
<td>.0408</td>
<td>.120</td>
<td>.952</td>
<td></td>
<td></td>
<td>.329</td>
<td>1.041</td>
<td>.961</td>
<td>1.127</td>
</tr>
</tbody>
</table>

Dependent Variable: IU_post
Model: (Intercept), Group, IU_pre

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

Table 2A

Negative Binomial (Constructed Dialogs) Counter-Alternative using pretest as covariate
### Table 3A

#### Negative Binomial (Constructed Dialogs) Counter-Critique using pretest as covariate

<table>
<thead>
<tr>
<th>Parameter</th>
<th>B</th>
<th>Std. Error</th>
<th>95% Wald Confidence Interval</th>
<th>Hypothesis Test</th>
<th>95% Wald Confidence Interval for Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>.435</td>
<td>.2936</td>
<td>-.140 to 1.011</td>
<td>2.200</td>
<td>.138 to 1.546</td>
</tr>
<tr>
<td>[Group=1.00]</td>
<td>-.140</td>
<td>.2971</td>
<td>-.722 to .442</td>
<td>.222</td>
<td>.637 to .869</td>
</tr>
<tr>
<td>[Group=2.00]</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CA_pre</td>
<td>.174</td>
<td>.0903</td>
<td>-.003 to .351</td>
<td>3.713</td>
<td>.054 to 1.190</td>
</tr>
<tr>
<td>(Scale)</td>
<td></td>
<td>1b</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Negative binomial)</td>
<td>1b</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dependent Variable: CA_post
Model: (Intercept), Group, CA_pre

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

### Table 4A

<table>
<thead>
<tr>
<th>Parameter</th>
<th>B</th>
<th>Std. Error</th>
<th>95% Wald Confidence Interval</th>
<th>Hypothesis Test</th>
<th>95% Wald Confidence Interval for Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-.971</td>
<td>.3581</td>
<td>-.672 to -.269</td>
<td>7.347</td>
<td>.007 to .379</td>
</tr>
<tr>
<td>[Group=1.00]</td>
<td>-.292</td>
<td>.4057</td>
<td>-.1087 to .503</td>
<td>.517</td>
<td>.472 to .747</td>
</tr>
<tr>
<td>[Group=2.00]</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CC_pre</td>
<td>.329</td>
<td>.1604</td>
<td>.014 to .643</td>
<td>4.202</td>
<td>.040 to 1.389</td>
</tr>
<tr>
<td>(Scale)</td>
<td></td>
<td>1b</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Negative binomial)</td>
<td>1b</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dependent Variable: CC_post
Model: (Intercept), Group, CC_pre

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.
**Negative Binomial (Constructed Dialogs) Counter-Undermine using pretest as covariate**

### Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>B</th>
<th>Std. Error</th>
<th>95% Wald Confidence Interval</th>
<th>95% Wald Confidence Interval for Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>.339</td>
<td>.2763</td>
<td>-.203</td>
<td>.881</td>
</tr>
<tr>
<td>[Group=1.00]</td>
<td>-2.94</td>
<td>.3156</td>
<td>-.912</td>
<td>.325</td>
</tr>
<tr>
<td>[Group=2.00]</td>
<td>0*</td>
<td>. .</td>
<td>. .</td>
<td>. .</td>
</tr>
<tr>
<td>CU_pre</td>
<td>.089</td>
<td>.0915</td>
<td>-.090</td>
<td>.269</td>
</tr>
<tr>
<td>(Scale)</td>
<td>1b</td>
<td>. .</td>
<td>. .</td>
<td>. .</td>
</tr>
<tr>
<td>(Negative binomial)</td>
<td>1b</td>
<td>. .</td>
<td>. .</td>
<td>. .</td>
</tr>
</tbody>
</table>

Dependent Variable: CU_post  
Model: (Intercept), Group, CU_pre

a. Set to zero because this parameter is redundant.  
b. Fixed at the displayed value.

---

**Table 5A**

**Negative Binomial Evidence (Constructed Dialogs) using pretest as covariate**

### Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>B</th>
<th>Std. Error</th>
<th>95% Wald Confidence Interval</th>
<th>95% Wald Confidence Interval for Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>.294</td>
<td>.2753</td>
<td>-.246</td>
<td>.833</td>
</tr>
<tr>
<td>[Group=1.00]</td>
<td>-.287</td>
<td>.3075</td>
<td>-.390</td>
<td>.316</td>
</tr>
<tr>
<td>[Group=2.00]</td>
<td>0*</td>
<td>. .</td>
<td>. .</td>
<td>. .</td>
</tr>
<tr>
<td>Ev_pre</td>
<td>.295</td>
<td>.0900</td>
<td>.119</td>
<td>.472</td>
</tr>
<tr>
<td>(Scale)</td>
<td>1b</td>
<td>. .</td>
<td>. .</td>
<td>. .</td>
</tr>
<tr>
<td>(Negative binomial)</td>
<td>1b</td>
<td>. .</td>
<td>. .</td>
<td>. .</td>
</tr>
</tbody>
</table>

Dependent Variable: Ev_post  
Model: (Intercept), Group, Ev_pre

a. Set to zero because this parameter is redundant.  
b. Fixed at the displayed value.
Table 6A

Negative Binomial (Constructed Dialogs) However Statement Units using pretest as covariate

<table>
<thead>
<tr>
<th>Parameter</th>
<th>B</th>
<th>Std. Error</th>
<th>Lower</th>
<th>Upper</th>
<th>Wald Chi-Square</th>
<th>df</th>
<th>Sig.</th>
<th>Exp(B)</th>
<th>95% Wald Confidence Interval for Exp(B)</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-.301</td>
<td>.1172</td>
<td>-.923</td>
<td>.321</td>
<td>.901</td>
<td>1</td>
<td>.343</td>
<td>.740</td>
<td>.397</td>
<td>1.378</td>
<td></td>
</tr>
<tr>
<td>[Group=1.00]</td>
<td>-.350</td>
<td>.3733</td>
<td>-1.081</td>
<td>.382</td>
<td>.877</td>
<td>1</td>
<td>.349</td>
<td>.705</td>
<td>.339</td>
<td>1.465</td>
<td></td>
</tr>
<tr>
<td>[Group=2.00]</td>
<td>0a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS_pre</td>
<td>.079</td>
<td>.1236</td>
<td>-.163</td>
<td>.321</td>
<td>.406</td>
<td>1</td>
<td>.524</td>
<td>1.082</td>
<td>.849</td>
<td>1.378</td>
<td></td>
</tr>
<tr>
<td>(Scale)</td>
<td>1b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Negative binomial)</td>
<td>1b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dependent Variable: HS_pre
Model: (Intercept), Group, HS_pre

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.
Table 7A

Comparison Group Summary (Present study)

<table>
<thead>
<tr>
<th>Codes</th>
<th>Comparison Group /Pre-intervention ever-used</th>
<th>Mean &amp; SD</th>
<th>Comparison Group /Post-intervention ever-used</th>
<th>Mean</th>
<th>McNemar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counter-Alternative</td>
<td>74%</td>
<td>1.63 (SD 1.57)</td>
<td>87%</td>
<td>1.8 (SD 1.51)</td>
<td>NS</td>
</tr>
<tr>
<td>Counter-Critique (Direct-Counter)</td>
<td>77%</td>
<td>1.23 (SD 1.06)</td>
<td>37%</td>
<td>0.46 (SD 0.70)</td>
<td>NS</td>
</tr>
<tr>
<td>Counter-Undermine (Direct-Counter)</td>
<td>60%</td>
<td>1.29 (SD 1.45)</td>
<td>69%</td>
<td>1.17 (SD 1.27)</td>
<td>NS</td>
</tr>
<tr>
<td>Disagree</td>
<td>9%</td>
<td>0.09 (SD 0.28)</td>
<td>11%</td>
<td>0.11 (SD 0.32)</td>
<td>NS</td>
</tr>
<tr>
<td>Meta</td>
<td>23%</td>
<td>0.46 (SD 0.92)</td>
<td>23%</td>
<td>0.37 (SD 0.77)</td>
<td>NS</td>
</tr>
<tr>
<td>Questioning</td>
<td>9%</td>
<td>0.09 (SD 0.28)</td>
<td>14%</td>
<td>0.14 (SD 0.36)</td>
<td>NS</td>
</tr>
<tr>
<td>Concession</td>
<td>57%</td>
<td>1.06 (SD 1.16)</td>
<td>60%</td>
<td>0.89 (SD 0.90)</td>
<td>NS</td>
</tr>
<tr>
<td>However statements</td>
<td>60%</td>
<td>1.03 (SD 1.01)</td>
<td>37%</td>
<td>0.57 (SD 1.01)</td>
<td>NS</td>
</tr>
<tr>
<td>Evidence</td>
<td>49%</td>
<td>0.94 (SD 1.39)</td>
<td>60%</td>
<td>1.51 (SD 2.01)</td>
<td>NS</td>
</tr>
<tr>
<td>Idea Units Total</td>
<td>262</td>
<td>7.49 (SD 2.80)</td>
<td>237</td>
<td>6.77 (SD 2.76)</td>
<td>NS</td>
</tr>
</tbody>
</table>
### Table 4.
Present Study: Constructed Dialogs intervention group/ pretest.

<table>
<thead>
<tr>
<th>Codes</th>
<th>% Ever-used in initial CD</th>
<th>Mean Number of Uses + SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counter-Alternative</td>
<td>85%</td>
<td>2.18 (1.39)</td>
</tr>
<tr>
<td>Counter-Critique</td>
<td>67%</td>
<td>1.23 (1.23)</td>
</tr>
<tr>
<td>Counter-Undermine</td>
<td>74%</td>
<td>1.85 (1.69)</td>
</tr>
<tr>
<td>Disagree</td>
<td>26%</td>
<td>0.28 (0.51)</td>
</tr>
<tr>
<td>Meta</td>
<td>21%</td>
<td>0.33 (0.70)</td>
</tr>
<tr>
<td>Questioning</td>
<td>8%</td>
<td>0.13 (0.47)</td>
</tr>
<tr>
<td>Concession</td>
<td>64%</td>
<td>1.54 (1.79)</td>
</tr>
<tr>
<td>However statements</td>
<td>69%</td>
<td>1.67 (1.71)</td>
</tr>
<tr>
<td>Evidence</td>
<td>79%</td>
<td>2.05 (1.85)</td>
</tr>
</tbody>
</table>

### Table 5.
Kuo & Kuhn, 2024. A-only group/ pretest.

<table>
<thead>
<tr>
<th>Codes</th>
<th>% Ever-used in initial CD</th>
<th>Mean Number of Uses + SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counter-Alternative</td>
<td>100%</td>
<td>3.87 (1.71)</td>
</tr>
<tr>
<td>Counter-Critique</td>
<td>39%</td>
<td>0.43 (0.59)</td>
</tr>
<tr>
<td>Counter-Undermine</td>
<td>17%</td>
<td>0.17 (0.39)</td>
</tr>
<tr>
<td>Disagree</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Meta</td>
<td>2%</td>
<td>0</td>
</tr>
<tr>
<td>Questioning</td>
<td>9%</td>
<td>0.09 (0.29)</td>
</tr>
<tr>
<td>Concession</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>However statements</td>
<td>9%</td>
<td>0.13 (0.34)</td>
</tr>
<tr>
<td>Evidence</td>
<td>91%</td>
<td>2.09 (1.62)</td>
</tr>
</tbody>
</table>