International Journal of Educational Research 126 (2024) 102375

Contents lists available at ScienceDirect

International Journal of Educational Research

journal homepage: www.elsevier.com/locate/ijedures

The effects of discourse goals on written arguments in elementary school students

Yuchen Shi

Institute of Curriculum and Instruction, Faculty of Education, East China Normal University, Zhongshan North Road #3663, Shanghai, China

ARTICLE INFO

Keywords:
Argumentative dialogue
Argumentative writing
Chinese elementary school students
Counterargument
Discourse Goal
Evidence

ABSTRACT

This study investigates the impact of discourse goals on argumentative writing among Chinese elementary school students. Eight fourth-grade classes were assigned to one of three conditions—Persuasion Dialogue (PD), Deliberation Dialogue (DD) and Control—to discuss four controversial topics designed for the Morality and Law subject. Before writing on the intervention topics, PD and DD students engaged in direct peer-to-peer argumentative written dialogues, while Control students addressed the topics and textbook contents in traditional teacher-led whole-class discussions. In essays on multiple intervention topics and a post-assessment topic, PD and DD students outperformed Control students in considering an alternative viewpoint. PD students, however, showed an advantage over DD students and Control students, in rebutting a counterargument and in attributing these counterarguments to the preceding dialogues, as well as in employing evidence to support claims. DD students, in contrast, were more likely to employ a dismissal strategy over a refutation strategy. Strengths of the argumentative discourse goal of persuasion over deliberation for elementary school students are discussed.

1. Introduction

Given the foundational role of argumentation in higher-order thinking (Kuhn, 1991; Mercier & Sperber, 2011; Moshman, 2011) and learning (Andriessen et al., 2003; Asterhan & Schwarz, 2016), educational researchers and practitioners across subject areas have made increasing efforts to foster students' argument skills (e.g., Asterhan & Schwarz, 2007; De La Paz et al., 2012; Newell et al., 2011; Rapanta & Felton, 2022; Zohar & Nemet, 2002). However, argumentative writing remains a challenge for students of all ages. Indeed, a growing line of research indicates that most students have not attained adequate skills to support successful argumentative writing (e.g., Applebee et al., 1990; Ferretti et al., 2000; Graham & Perin, 2007; Wolfe, 2011).

The present study follows earlier authors in recognizing writing as a social practice (Applebee et al., 2003; Billig, 1987; VanDerHeide & Newell, 2013). An individual, when writing, needs to engage in a solitary dialogue with a hypothetical other, a "missing interlocutor" (Graff, 2003). Research indicates that holding argumentative dialogues with others helps students internalize the argument-counterargument structure and promotes their consideration of alternative viewpoints in writing (e.g., Kuhn et al., 2016; Reznitskaya et al., 2001). Research further indicates that in contrast to the persuasion goal of dialogue (i.e., to persuade an opponent), the deliberation goal of dialogue (i.e., to reach consensus with an opponent) enabled students to generate stronger and more integrated arguments, both in dialogues and in writing (Asterhan & Babichenko, 2015; Asterhan & Schwarz, 2016; Felton et al., 2015b; Villarroel et al., 2016).

E-mail address: ys2786@tc.columbia.edu.

https://doi.org/10.1016/j.ijer.2024.102375
Received 11 July 2023; Received in revised form 17 April 2024; Accepted 1 May 2024
Available online 13 May 2024

However, existing research along this line has primarily examined the performance of secondary school students or young adults, and there is a lack of research investigating the effects of discourse goals on the argumentative writing of elementary school students. On one hand, it is possible that the discourse goal of deliberation (e.g., "to build knowledge") is more abstract for elementary school students to understand, they may benefit while still young from first engaging in persuasion dialogues that they are more familiar with, which helps develop skills of counterargument and rebuttal. On the other hand, the advantages associated with engaging in deliberation dialogues reported among older participants in prior studies might be applicable to younger participants as well.

In the present study, eight equivalent classes of fourth-grade Chinese elementary school students were assigned to three conditions—Deliberation Dialogue, Persuasion Dialogue and Control. Essays on each of the four intervention topics, as well as on a non-discourse topic at pre- and post-assessments, were examined to investigate whether engaging in dialogues with differing goals led to differences in students' argumentative writing performance.

1.1. Weaknesses in students' argumentative writing

Although children tend to develop oral argumentative capabilities at a very young age, their abilities to successfully compose argumentative texts tend to develop much later (Golder & Coirier, 1994). Several weaknesses in students' argumentative writing have been identified, and prominent among them were their difficulty in drawing on evidence to justify claims (Berland & Reiser, 2011; Clark & Sampson, 2007; Iordanou & Constantinou, 2015), as well as in weighing arguments and counterarguments within a framework of alternatives (Kuhn, 1991; Leitão, 2000, 2003; Pollack, 1987).

Students often struggle with coordinating claims with evidence (Brem & Rips, 2000; Kuhn, 1991), which has been recognized as a core component of arguments by many theorists (Baker, 2003; Toulmin, 1958; Walton & Zhang, 2013). Research shows that students tend to fail to cite sufficient data to support claims (Sandoval & Millwood, 2005), or, when they do recognize the need to refer to data, they frequently do not include adequate warrants or backings to specify how specific data related to particular claims (Bell & Linn, 2000; Ryu & Sandoval, 2012; Sandoval & Millwood, 2005). This frequently observed difficulty in claim-evidence coordination is connected to a long-standing line of research reporting on individuals' preference for explanation (of mechanism) over evidence in justifying claims (Brem & Rips, 2000; Kuhn, 1991). Students' failure to recognize evidence as distinct from claim and bearing on it precludes adhering to a theory-evidence coordination model in which multiple alternatives are considered, and a claim that has the most consistent and least inconsistent evidence associated with it is the alternative that is chosen.

Relatedly, although the central role of counterargument in argumentation has been widely recognized at theoretical and empirical levels in argumentation studies (Erduran et al., 2004; Toulmin, 1958; van Eemeren et al., 1987; Walton, 1989), students' rhetorical arguments in writing are often one-sided. As noted in Nussbaum and Schraw (2007), many researchers view the construction of counterarguments as central to sophisticated, open-minded thinking (Bakhtin, 1981; Baron, 1988; Vygotsky, 1962), and a key indicator of an individual's reasoning competence (Kuhn, 1991; Perkins et al., 1991). However, students often fail to address counterarguments to their positions in writing, let alone integrating counterarguments with the rest of their arguments.

From a cognitive perspective, students' failure to include counterarguments in writing might be related to 'my-side' bias (Felton et al., 2015a; Perkins et al., 1991; Stanovich & West, 2007; Wolfe, 2012), a concept originally defined by Perkins (1985) as one's tendency to ignore the evidence against one's favored position, probably due to self-centeredness or egocentrism. Young students' innate my-side bias (Mercier & Sperber, 2011) and an underdeveloped argument schema (Anderson & Pearson, 1984; Reznitskaya et al., 2012; Wolfe et al., 2009) makes it hard for them, or failing to recognize the need, to extract or "decouple" from their viewpoints (Kuhn, 2022; Leitão, 2003). From a dialogical perspective (Vygotsky, 1978), writing is a social practice (Applebee, 1996; VanDerHeide & Newell, 2013; Kuhn et al., 2016) in which a writer constructs a solitary discourse with a hypothetical other (Shi, 2020a; Zavala & Kuhn, 2017). This recognition of a 'virtual other' in the construction of arguments resonates with contemporary view that acknowledges the dialogical dimension of individuals' thinking (Bakhtin, 1981; Mead, 1934/1972; Wertsch, 1991). However, generating dialogic discourse from an individual's mind—that is, predicting, acknowledging, and responding to potential objections—demands specific linguistic skills and the complex management of textual devices (Coirier, 1996; Feilke, 1996; Golder & Coirier, 1996).

At the same time, for successful argumentative writing, students need to integrate counterarguments with other arguments, rather than leaving counterarguments unaddressed. Nussbaum and Schraw (2007) identified several strategies for the argument-counterargument integration, including *refutation,* in which one tries to refute the counterargument by showing it is somehow flawed; *weighing,* in which one considers both sides' advantages and disadvantages and shows that the benefits outweigh the negative consequences; and *synthesis,* in which one arrives at a final standpoint that circumvents or eliminates a problem. Casado--Ledesma (2021) argued that although *refutation* allows the problem space to be explored, *weighing* or *synthesis* are more integrative strategies that encourage two-sided reasoning (Felton et al., 2009; Nussbaum & Schraw, 2007). Still, Nussbaum (2008) pointed out that the most appropriate integration strategy was the one that met the goal of the argumentation task - if the goal was to persuade a reader, then *refutation* would be an appropriate strategy to respond to a potential counterargument.

1.2. The effect of discourse goals on argumentative writing

Existing research examining the influence of discourse goals on argumentative writing largely falls into two strands. The first treats writing as a form of solitary discourse and examines how writing goals affect essay quality (e.g., Ferretti et al., 2000; Nussbaum & Kardash, 2005); the second focuses on peer discourse and examines how engaging in dialogues with differential goals affects subsequent writing performance. Studies in the first strand report that persuasion goals undermined the quality of written arguments, whereas providing explicit subgoals about the elements of an argumentative essay helped elementary school (Ferretti et al., 2000) and

college students (Nussbaum & Kardarsh, 2005) write more persuasive essays that included more argumentative elements, such as counterarguments and rebuttals.

Studies in the second strand follow the socio-cultural approach in investigating the influence of dialogues on writing. According to Mercier and Sperber (2011), argumentative discourse serves two fundamental communicative functions: to convince others to accept an argument (i.e., to persuade) and to assess arguments posed by others (i.e., to deliberate). While persuasion dialogue is an adversarial exchange in which speakers advance competing claims to convince others to accept or adopt their position, deliberation dialogue is a collaborative exchange in which speakers holding incompatible claims seek to resolve their differences of opinion to arrive at a consensual decision (Walton & Krabbe, 1995). Research using adolescents or young adults as participants converged to demonstrate the advantages of deliberation dialogues over persuasion dialogues, both in content learning and in writing performance (Felton et al., 2015a, 2015b; Golanics & Nussbaum, 2008; Villaroel et al., 2016).

In an experiment by Asterhan and Babichenko (2015), undergraduates engaged in computer-mediated interactions with a confederate on their understanding of a scientific concept. The peer confederate's verbal behavior was scripted to evoke argumentative discourse while controlling exposure to conceptual content and the type of dialogue moves but differing in argumentative discourse style (i.e., disputative or deliberative). Results showed that learners who participated in the deliberative discourse style condition outperformed those in the disputative condition on individual learning scores.

Felton, Crowell and Liu (2015a) asked college students to engage in argumentative dialogues before writing essays supporting their opinions on capital punishment. Students who argued to reach consensus were more likely to cite position-incongruent arguments initially raised by their dialogue partners, and to reconcile valid position-incongruent arguments with their global position through the synthesis strategy. In contrast, students who argued to persuade were less likely to cite position-incongruent arguments in their writing and, when they did, tended merely to list or refute them.

Villarroel and colleagues (2016) grouped pre-service teachers into disagreeing dyads and asked them to either argue to persuade or argue to reach consensus. As they conducted argumentative dialogues, participants were presented with three types of graphical evidence—supporting their view, challenging their view, or ambiguous. When writing argumentative essays at post-test, consensus condition participants were more likely to interpret disconfirming evidence correctly and refer to their dialogues more frequently. This finding suggests that holding dialogues with consensus goals helped pre-service teachers reduce confirmation bias in writing.

1.3. The present study

As illustrated above, prior research demonstrating the superiority of deliberation over persuasion dialogues in supporting students' argumentative writing has primarily examined the adolescent or adult populations. The question regarding the comparative effects of the two discourse goals among students in the elementary grades remains unexplored. Children might benefit more from persuasion than deliberation, as they tend to have significant experience with arguing to persuade others or to win a dispute in their daily lives (Golder, 1992; Golder & Coirier, 1994; Stein & Bernas, 1999), and a paucity of experience with engaging in more collaborative forms of arguing (Felton et al., 2009; Lind et al., 2023; Stein & Miller, 1991). Deeper learning of argument skills might have taken place when students engage in the type of argumentative discourse they are more familiar with. An alternative hypothesis was that because of children's lack of experience engaging in deliberation in everyday life, they might derive more benefits from deliberative practices that hold the potential to diminish defense motivation and prompt individuals to reconcile conflicting perspectives (Villaroel et al., 2016).

To investigate these competing hypotheses, the present study employed a pre- and post-test quasi-experimental design and assigned eight fourth-grade classes from a Chinese elementary school into three conditions: Persuasion dialogue (PD) condition, Deliberation dialogue (DD) condition and Control condition. While PD and DD students engaged in argumentative peer dialogues on a series of controversial issues designed based on textbook contents (Shi et al., 2021, 2024), Control students discussed the same set of topics and their related textbook contents in a whole-class, teacher-led format for the same duration. In line with prior research that has consistently demonstrated the developmental trajectory of argument skills (Berland & McNeill, 2010; Crowell & Kuhn, 2014; Hemberger et al., 2017; Osborne et al., 2016; Shi, 2019) over time, the present research designed an extended intervention to repeatedly engage PD and DD students in argumentative dialogues to foster and allow close examination of changes over time. The design of the study allows us to investigate two research questions:

(1) For essays on each intervention topic, are there between-condition differences in the effect of argumentative discourse goals on generating and integrating arguments inconsistent with one's position, as well as in drawing on evidence to support arguments?
(2) For essays on a non-discourse topic at post-assessment, are such differences maintained between these two intervention conditions and in comparison to a non-intervention control condition?

2. Methods

2.1. Participants

Eight equivalent fourth-grade classes from a public elementary school in a large city in eastern China participated in the present study. The school was among the top-tier in the neighborhood, primarily serving middle- to upper-middle-class Chinese families whose residential address was nearby. The school administrators informed us that the eight fourth-grade classes had been composed so as to be equivalent with respect to students' gender, academic ability and achievement, and social characteristics.

Six months prior to the study, the research team contacted the school principal to indicate the intent to conduct the study. The

school administrators then assigned five classes to participate in the study and from them, the research team randomly selected three to serve in the Persuasion Dialogue (PD) condition ($n = 100$, 45 girls and 55 boys) and two to serve in the Deliberation Dialogue (DD) condition ($n = 68$, 30 girls and 38 boys). The remaining three classes served in the Control condition ($n = 99$, 43 girls and 56 boys). The students were aged from nine years, seven months to ten years, four months at the start of the intervention; they all spoke and wrote in Mandarin Chinese, their native language.

2.2. Design

The study was quasi-experimental involving repeated measures with a pre-and post-test control group design. All students wrote essays on each of the four intervention topics and a non-discourse topic at pre- and post-assessments. Students differed, however, in whether and how they participated in the dialogue component of the intervention. Specifically, for each intervention topic, (1) students in the PD condition engaged in two sessions of argumentative dialogues via a written form with a different disagreeing peer in each session, with the goal of persuading the peer to adopt their position; (2) students in the DD condition also engaged in two sessions of argumentative dialogues in the same format, but their goal was to reach consensus or a mutually agreed-upon solution through deliberation; (3) students in the Control condition did not engage in direct peer dialogues but were taught in a business-as-usual, whole-class setting, where classroom teachers covered the intervention topics in connection with textbook contents following the Initiation-Response-Evaluation (IRE) discourse pattern (Mehan, 1979; Sinclair & Coulthard, 1975). To ensure treatment fidelity, the research team observed, recorded and took field notes during all the sessions in the three conditions.

2.3. Procedure

2.3.1. Pre- and post-assessments

One week before the start of the intervention, an individual pre-assessment was administered to all the students in the form of a written task in a whole-class setting. Students were given the entire class period (35 min) to complete the task. The same assessment was re-administered during the week following the intervention. Specifically, students were given the prompt shown in Table 1.

Along with the essay prompt, three reports based on authentic sources were made available to each student as an information sheet. These reports contained conflicting evidence on the issue, with some supporting the benefits of obtaining a college degree and others the advantages of receiving vocational training.

2.3.2. Design of controversial topics

Controversial topics designed and applied in Shi et al. (2024) were re-used in the present study. In Shi et al. (2024), the researchers and participating teachers collaboratively designed four controversial topics (shown in Table 2) based on the fourth-grade textbook in Morality and Law. The textbook was used nationally across China. Specifically, four textbook units were selected and a controversy was designed by reorganizing one unit of textbook contents around it. The team was guided by three principles in designing these controversial topics: first, the topic had to serve as a "unit organizer" capturing the unit's key contents; second, the topic had to be ill-structured to allow extended exploration of alternative perspectives; third, the issue had to be connected to real life so that students could invoke and apply their authentic life experiences.

2.3.3. Intervention procedure

For each intervention topic, students were encouraged first to take a side in an opinion poll; each student was then paired with an opposing-side peer to conduct one-on-one argumentative dialogue via "pass-the-pad," in which one student wrote on a pre-designed sheet and physically pass it to their designated opponent. Each student was paired with a different opponent for the two dialogue sessions. At the start of each session, the teachers reminded students of their discourse goal, with PD students receiving the reminder to "Try to convince your opponent to switch position" and DD students receiving the reminder to "Try to work with your opponent to reach an agreement or find a solution to the problem." The same directives were also printed at the top of each sheet used for written dialogues. Evidence was not provided in Topic 1 as students were still gaining familiarity with the dialogue procedure. Topic 1 thus served as a baseline to determine the extent to which students drew on their personal knowledge to support claims when no external information was provided. For Topics 2, 3 and 4, topic-related evidence, in the Question-and-Answer format, was provided to each student as they engaged in dialogues. All the evidence provided were based on data or information that were publicly available and verifiable. For each topic, twelve pieces of evidence were provided, with six pieces supporting the pro side and six the con side. Sample evidence is provided in Table 3.

For each intervention topic, following the two dialogue sessions (PD and DD conditions) or the teacher-led discussion sessions

Table 1
Essay topic at pre- and post-assessments.

Topic scenario: Yao was a 4th-grade student and was extremely fond of cooking. In her spare time, she liked to help her parents with cooking. Her goal was to get into a culinary school following graduation from middle school so she could become a cook in the future. However, her parents wanted her to get into a prestigious college so that she could find a more decent job. However, this wasn't what Yao wanted, which made her quite distressed and uncertain of what to do. **Question:** Do you support Yao or her parents? Please write a letter to Yao to make clear your position. Support your position with reasons and evidence. There is no word limit.

Table 2
List of intervention topics.

Topic Number	Topic statement
1	**Scenario:** Lan and Yao were two 4th-grade students who were desk mates. During class breaks, Lan liked to play with Yao's new stationeries; sometimes, he accidentally broke them. Over time, Yao felt increasingly offended that Lan always played with others' personal belongings without permission. She told Lan that his behavior wasn't acceptable. However, although Lan apologized, promising to seek Yao's permission first, he never adhered to his promise. **Question:** Should Yao continue to forgive Lan? Please write a letter to Yao.
2	**Scenario:** One Saturday morning, Lan's mother was unexpectedly called to take an extra shift at work. Before going to work, she needed to send Lan to his calligraphy class. She had no time to cook lunch and decided to order delivery foods. However, the delivery guy had been late for 20 min, making Lan and his mother quite anxious as they might have to go to work or class with empty stomachs. They called the delivery guy several times and each time, he responded by saying he was about to arrive. Lan's mother was quite angry and decided to give the guy a very negative rating. **Question:** Should Lan persuade his mother not to give a negative rating? Please write a letter to Lan.
3	**Scenario:** Yao was a 4th-grade student with many electronic devices in her household, including TV, mobile phones, computer, iPad, etc. Yao liked to use these devices for academic and entertainment purposes. However, since last semester, Yao's school grades started to fall behind, and she began to develop myopia. Her parents wanted to limit her "screen time" and only allowed her to do school work on these devices. However, Yao believed these devices were important to her as she could obtain tremendous resources for learning. Besides, she could use some entertainment and relaxation time, which was very necessary for life. **Question:** Do you support Yao or her parents? Please write a letter to Yao.
4	**Scenario:** Lan visited his hometown with his father every summer vacation. The forest resources in his hometown were very rich, but the local economy was quite poor. Residents made minimal earnings, and many young people had no choice but to leave their hometown to search for employment opportunities in the urban area. Lan learned from his father that to change this situation, the local government wanted to make use of the rich forest resources to build a timber mill, which would not only provide plenty of job opportunities for the young so that they don't have to leave their hometown but also increase government revenues, which could be used to improve the local construction. However, the timber mill involves cutting down trees and taking up arable land, not to mention industrial waste that might damage the environment and threaten residents' health. **Question:** Do you think the local government should build the timber mill? Please write a letter to Lan.

(Control condition), all students completed an individual argumentative writing task in a whole-class setting. The list of evidence provided to PD and DD students during argumentative dialogues was made available to each student. The writing prompt for each topic was as follows, "Please write a letter to your friend involved in the story. You need to make your position as clear as possible and support it with reasons and evidence. There is no word limit."

2.3.4. Analytic scheme for argumentative essays

The essays collected and analyzed for the present study included intervention essays and pre- and post-assessment essays. Each essay was first segmented into idea units, with a unit defined as a claim with supporting arguments or evidence. Fig. 1 provides a diagrammatic representation of the coding scheme, which consisted of four steps. In the first step, we analyzed the argumentative function of each idea unit, which was categorized as position-congruent, position-incongruent, or other units (Shi, 2019). Position-congruent units worked in one's favor as it supported one's position or weakened the opposing position; position-incongruent

Table 3
Sample evidence in the Question-and-Answer format for intervention topics.

Topic No.	Pro-position	Con position
2	**Q: Can consumers safeguard their rights and interests in the face of overtime delivery?** A: Absolutely. China enacted the *Consumer Protection Law of the People's Republic of China* in 1993, and it had undergone several rounds of revision. The law stipulates that when consumers purchase, use goods, or receive services, their rights and interests are protected by this law.	**Q: Is it dangerous to deliver takeaway?** A: Yes. In the first half of the year 2019, there were 325 traffic accidents involving express delivery and takeaway industry in Shanghai city, resulting in 5 deaths and 324 injuries. Statistics in 2020 showed that on average, one delivery guy was killed or seriously injured in traffic accidents every 2.5 days.
3	**Q: Do electronic devices make communication between people more convenient?** A: Yes, data released by Tencent in 2020 shows that there are as many as 1.21 billion WeChat users in China. Research from Nanyang Technological University in Singapore shows that the use of social networks has increased significantly during COVID-19. People can stay in touch and exchange information even if they are isolated at home.	**Q: Could using social media lead to psychological problems?** A: It's possible. A study conducted by the University of Pennsylvania in 2018 showed that using social media, such as WeChat Moments, might increase users' feeling of depression and loneliness. This may be because everyone likes to show their best on social media, and in comparing ourselves to others, we couldn't help but feel lonely, anxious or depressed.
4	**Q: What impact might a large number of young people going out to work have on left-behind children and the elderly?** A: Statistics released by the Ministry of Education in China in 2018 showed that the number of left-behind children in rural areas in 2017 was 15.5 million. These children cannot stay with their parents and lack appropriate education. In addition, as of the end of 2019, there were 130 million individuals over the age of 60 living in the rural area. Nearly half of them live alone for a long time, lacking proper care and are very lonely.	**Q: Are working in a timber mill hazardous to your health?** A: It is entirely possible that the processing of wooden boards produces toxic gases, with common ones including formaldehyde, benzene, etc. Long-term exposure may have adverse effects on the human respiratory tract, lungs, blood, etc.

Note. Please contact the author to obtain a full list of the evidence provided for each topic.

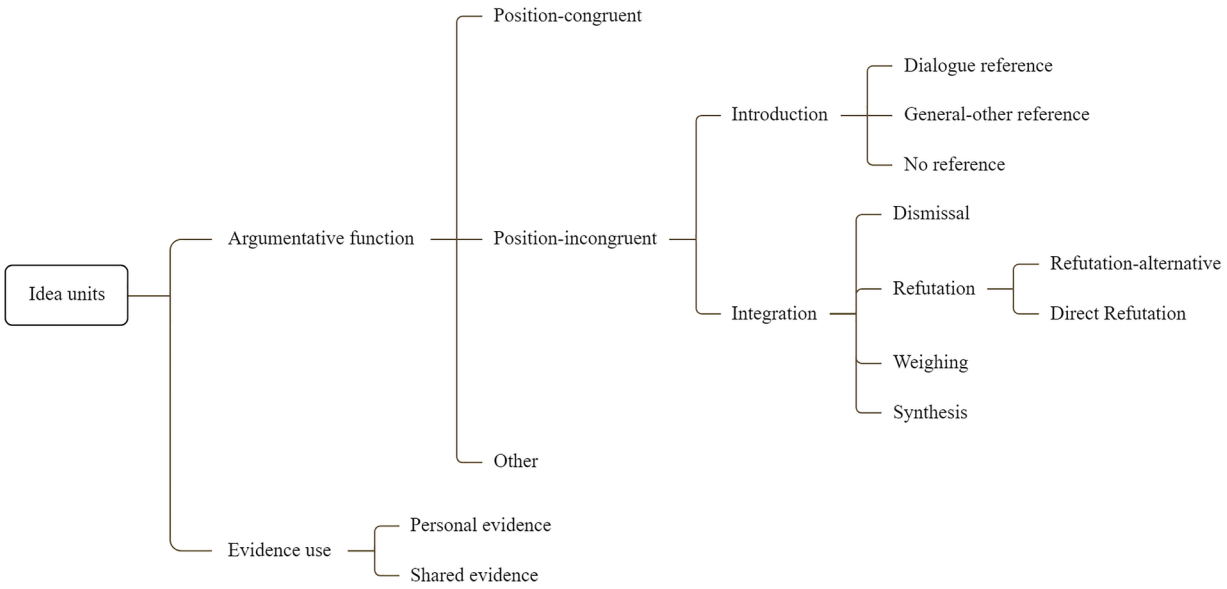

Fig. 1. Diagrammatic representation of the coding scheme.

units worked against one's favor as it weakened one's position or supported the opposing position; other units were neutral, without a discernible position. Blind to condition, the author and a colleague not involved in the present study coded a randomly selected 30 % of the essay dataset, achieving an inter-rater agreement of 90.91 % (Cohen's kappa = 0.873, $p < .0005$), in assigning an idea unit to one of three categories.

In the second step, we coded how each position-incongruent argument was introduced. To develop a coding scheme, the two coders collaboratively examined a portion of students' essays containing position-incongruent arguments in an open fashion, looking for recurrent theme. The two coders then discussed and summarized the themes into codes that were then applied to analyze more transcripts to check their applicability. New codes were added if new themes were identified. Eventually, the two coders identified three types of ways in which a position-incongruent argument was introduced: (1) *Dialogue reference*, in which the students explicitly referred to the preceding dialogues (e.g., "My partner told me that doing too much homework is the reason for developing myopia, not using cellphones."); (2) *General-other reference*, in which the students referred to others in general (e.g., "Some say using too much social media might lead to depression."); (3) *No reference*, in which the students neither referred to dialogues nor to others (e.g., "Using cellphones might hurt your eyes."). Working with 30 % of the dataset, the two coders independently assigned a position-incongruent argument into one of the three categories, achieving an inter-rater agreement of 97 % (Cohen's kappa = 0.88, $p < .0005$).

In the third step, we coded how each position-incongruent argument was integrated with the remaining arguments. Applying and modifying the integration strategy framework initially proposed by Nussbaum and Schraw (2007), we identified four types of integration: (1) *Dismissal,* in which a position-incongruent argument was listed but not attended to; (2) *Refutation,* in which one included a response that rebutted a counterargument; (3) *Weighing,* in which the evidence or reasons on one side was argued to outweigh those on the other; and (4) *Synthesis,* in which an in-between position or solution was developed to combine the merits of both sides. Verbatim examples for each integration strategy are provided in Table 4. Blind to condition, the two coders independently analyzed a randomly selected 30 % of the essay dataset, achieving an inter-rater agreement of 88.64 % (Cohen's kappa = 0.855, $p < .0005$), in assigning an

Table 4
Verbatim (translated) examples of each type of integration strategy (adapted from Nussbaum & Schraw, 2007).

Category of integration strategy	Subcategory	Example[1]
Dismissal		Some say parents should control their children, including how much time they use electronic devices.
Refutation[2]	Alternative	Some say there are a lot of resources at the school library, but I think you can have fun using the internet.
	Undermine	Although parents should protect their children, using electronic devices does not harm children.
	Critique	Although using electronic devices might cause myopia, one can do an easy surgery to correct it.
Weighing		If you use your iPhone less, your eyesight will get better. But you might be less happy. I think your happiness is more important.
Synthesis		Now you are fighting with your parents. I think you can talk to your parents so you both agree on how much time you could spend using these devices each week.

Note.
[1] Examples are derived from Topic 3 essays supporting Yao in using electronic devices for entertainment purposes.
[2] Refutation-undermine and Refutation-critique are combined as Direct Refutation.

integration strategy into one of four categories.

To further analyze how the *refutation* strategy was applied to weaken the original position-incongruent argument, we drew on the coding scheme originally developed in Kuhn and Papathomas (2017) and later modified in Shi (2020a). Specifically, in *Refutation-alternative*, the writer proposed an alternative claim to the original position-incongruent argument; it was considered as the weakest form of counterargument as it left the force of the original claim unaddressed. In *Refutation-critique*, the writer directly challenged the conclusion of the original claim and in *Refutation-undermine*, the writer directly weakened the premise of the original claim. *Refutation-critique* and *Refutation-undermine* were considered stronger as they directly attacked the original position-incongruent argument and therefore, we combined and referred to them as *Direct Refutation*. Verbatim examples for each type of *Refutation* are provided in Table 4. The two coders independently analyzed 30 % of the arguments coded as *Refutation* in the previous step, achieving an inter-rater agreement of 86 % (Cohen's kappa = 0.82, $p < .0005$), in assigning a *Refutation* move into one of three categories.

In the fourth step, we examined students' employment of evidence in the service of an argument. An idea unit was coded as evidence-based if it included reference to evidence, or non-evidence-based if it did not. Evidence referred to could be based on the information that had been provided (Shared evidence) or students' personal knowledge (Personal evidence). The two coders worked with 30 % of the dataset to assign an idea units into one of three categories—Personal evidence, Shared evidence, and non-evidence-based unit, achieving an agreement of 96 % (Cohen's kappa=0.93, $p < .0005$).

3. Results

The Results section is divided into three parts: Part One presents a statistical analysis of intervention essays; Part Two presents a statistical analysis of pre- and post-assessment essays; and Part Three presents case studies to qualitatively examine differences across conditions.

3.1. Intervention essays

3.1.1. Idea unit

The mean number of idea units (and standard deviations) for intervention essays and pre- and post-assessment essays is presented in Table 5. A Generalized Linear Model (GLM) with Poisson regression was conducted to examine differences in the mean number of idea units across conditions at each time point. The results indicated that none of the condition differences reached statistical significance, suggesting that at each topic, students from different conditions wrote essays of comparable length.

3.1.2. Position-congruent argument

Next, for each essay, we calculated its percentage of position-congruent arguments out of its total number of idea units, and the mean results are presented in Fig. 2. By calculating the percentage of position-congruent arguments, we controlled, to a certain extent, topic effects, which might be reflected in the number of idea units a student generated for that topic, with students likely to write more for easier topics and write fewer for more challenging ones. These percentages were subjected to a two-way mixed analysis of variance (ANOVA), with condition as a between-subjects factor, topic as a within-subjects factor, and percentage as an outcome variable. The Greenhouse-Geisser correction was applied since the assumption of sphericity was not met.

Of the four intervention topics, the percentage of position-congruent arguments showed a significant two-way interaction between condition and topic, F (5.423, 710.460) = 15.389, $p < .0005$, Partial $\eta^2 = 0.105$, $\varepsilon = 0.904$. Follow-up tests indicated no significant simple effects for condition at Topic 1 and Topic 2. However, significant simple effects for condition were detected at Topics 3 and 4. To adjust the level of statistical significance for multiple comparisons, a Bonferroni correction was made so that for each topic, acceptance of statistical significance was set at $p < .017$ as three comparisons were carried out.

At Topic 3, the percentage of position-congruent arguments was significantly lower in the Control condition compared to the PD condition ($M = -0.06$, $SE = 0.02$, $p < .0005$), as well as compared to the DD condition ($M = -0.05$, $SE = 0.02$, $p = .016$). No significant difference was found between the PD and DD conditions ($M = 0.02$, $SE = 0.02$, $p = .957$).

At Topic 4, the percentage of position-congruent arguments was significantly lower in the Control condition compared to the PD condition ($M = -0.13$, $SE = 0.02$, $p < .0005$), as well as compared to the DD condition ($M = -0.10$, $SE = 0.02$, $p < .0005$). No significant difference was found between the PD and DD conditions ($M = 0.03$, $SE = 0.02$, $p = .397$).

Table 5
Mean number of idea units (and standard deviations) in essays by condition and topic.

Condition	Pre-assessment	Topic 1	Topic 2	Topic 3	Topic 4	Post-assessment
PD condition	5.77	6.42	6.43	6.79	6.81	6.50
	(1.78)	(4.37)	(4.23)	(3.32)	(3.00)	(3.04)
DD condition	5.26	6.03	6.19	6.25	6.69	6.69
	(1.42)	(3.28)	(3.14)	(3.12)	(1.92)	(1.62)
Control condition	5.20	6.16	6.37	6.26	6.72	6.94
	(2.33)	(2.66)	(2.49)	(2.59)	(1.83)	(1.60)

Note. PD=Persuasion dialogue condition; DD=Deliberation dialogue condition.

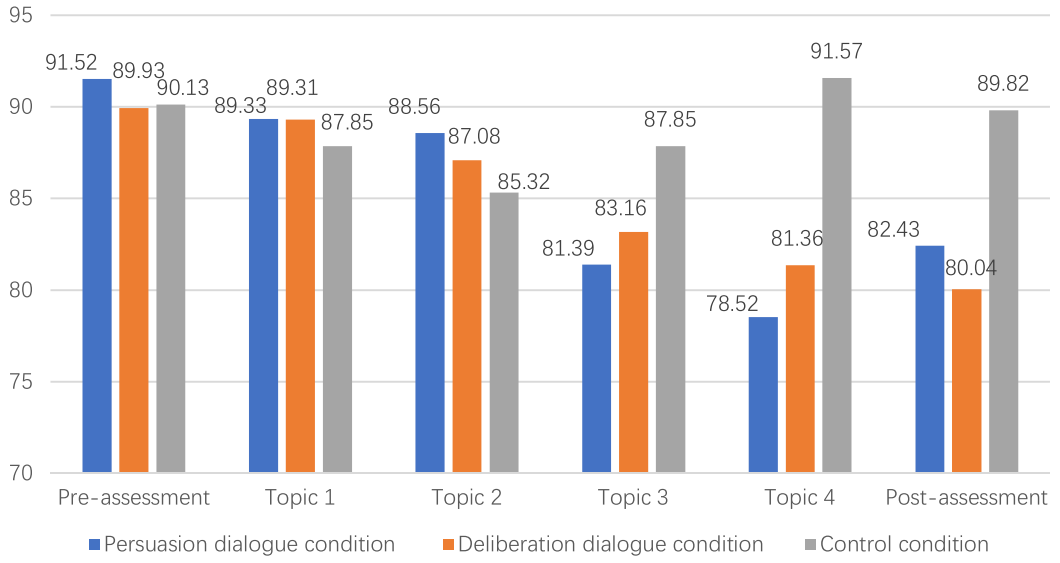

Fig. 2. Mean percentage of position-congruent arguments by condition and topic.

3.1.3. *Position-incongruent argument*

Next, for each essay, we calculated its percentage of position-incongruent arguments out of its total number of idea units, and the mean results are presented in Fig. 3. Of the four intervention topics, the percentage of position-incongruent arguments showed a significant two-way interaction between condition and topic, F (4.702, 616.018) = 6.257, $p < .0005$, Partial $\eta^2 = 0.046$, $\varepsilon = 0.784$. Follow-up tests indicated that except for Topic 1, significant simple effects for condition were detected at Topics 2, 3 and 4. Again, acceptance of statistical significance was set at $p < .017$.

At Topic 2, the percentage of position-incongruent arguments was lower in the Control condition compared to the DD condition, but the difference did not reach statistical significance when Bonferroni adjustment was applied ($M = -0.02$, $SE = 0.01$, $p = .018$). No significant difference was found between the PD and Control conditions ($M = 0.004$, $SE = 0.01$, $p = .712$), nor between the PD and DD conditions ($M = 0.01$, $SE = 0.01$, $p = .104$).

At Topic 3, the percentage of position-incongruent arguments was significantly lower in the Control condition compared to the PD condition ($M = -0.07$, $SE = 0.01$, $p < .0005$), as well as compared to the DD condition ($M = -0.08$, $SE = 0.02$, $p < .0005$). No significant difference was found between the PD and DD conditions ($M = 0.01$, $SE = 0.02$, $p = .902$).

At Topic 4, the percentage of position-incongruent arguments was significantly lower in the Control condition compared to the PD condition ($M = -0.05$, $SE = 0.01$, $p = .001$). No significant difference was found between the DD and Control conditions ($M = -0.03$,

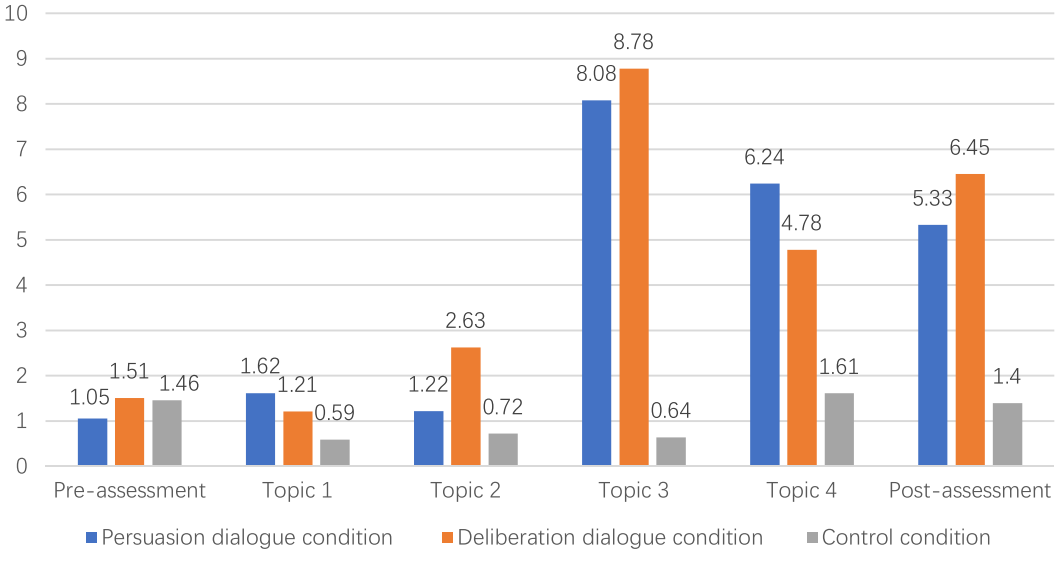

Fig. 3. Mean percentage of position-incongruent arguments by condition and topic.

SE = 0.01, $p = .061$), nor between the PD and DD conditions ($M = 0.01$, $SE = 0.01$, $p = .543$).

Next, to analyze the strategy students used to introduce a position-incongruent argument, we focused on the use of *Dialogue reference* statement, which reflected a direct influence of dialogues on the subsequent writing activity. For all intervention topics, zero percent of Control students included a *Dialogue reference* statement. Therefore, subsequent Fisher's exact test was carried out based on results from the PD and DD conditions. At Topic 1, 7 % of PD students and 5.9 % of DD students ever made a *Dialogue reference* statement, a non-significant difference ($p = 1.000$). At Topic 2, 6 % of PD students and 7.4 % of DD students ever made a *Dialogue reference* statement, a non-significant difference ($p = .758$). At Topic 3, 22 % of PD students and 8.8 % of DD students ever made a *Dialogue reference* statement, a significant difference ($p = .034$). At Topic 4, 15 % of PD students and 4.4 % of DD students ever made a *Dialogue reference* statement, a significant difference ($p = .040$). Taken together, in the last two intervention topics, even though PD and DD students made a comparable number of position-incongruent arguments, PD students were more likely than DD students to attribute the argument to the preceding dialogues.

3.1.4. Integration of position-incongruent argument

Next, we calculated the percentage of students who ever used a certain integration strategy to incorporate a position-incongruent argument (see Table 6). Since Control students rarely included position-incongruent arguments, we focused on PD and DD students for this part of the analysis. In addition, since more sophisticated integration strategies—*Weighing* and *Synthesis*—were rarely used, we combined them and calculated the number of students who used either. Applying Fisher's Exact test, our results indicated that at Topic 3 ($p = .027$) and Topic 4 ($p = .01$), a significantly higher percentage of DD than PD students ever employed *Dismissal*. At the same time, at Topic 3 ($p = .001$) and Topic 4 ($p = .002$), a significantly higher percentage of PD than DD students ever employed *Refutation*. For all intervention topics, the percentage of students who ever employed *Weighing* or *Synthesis* remained negligible and did not significantly differ across conditions.

We further differentiated between *Refutation-Alternative* and *Direct Refutation*. Focusing on *Direct Refutation*, the more successful strategy of the two, we calculated the percentage of *Direct Refutation* relative to the total number of *Refutation* for each topic. The results indicated that across the four intervention topics, the percentages of *Direct Refutation* ranged between 80 % and 85 % in the PD condition, compared to the range between 20 % and 25 % in the DD condition, suggesting that PD students were more successful in directly refuting a position-incongruent argument, as early as in Topic 1.

3.1.5. Generation of evidence-based argument

The mean percentages of evidence-based units out of the total number of ideas units are presented in Table 7. As shown in Table 7, the mean percentage of evidence-based units based on Personal evidence was around 10 %, which remained stable across conditions and topics. A two-way mixed ANOVA with condition as a between-subjects variable and topic as a within-subjects variable indicated that for Personal evidence, there was no significant interaction of condition and topic, F (5.145, 674.022) = 1.423, $p = .212$, Partial $\eta^2 = 0.011$, $\varepsilon = 0.858$. Follow-up tests of the main effects of condition and topic also did not reach significance.

For Shared evidence, we excluded Topic 1 and focused on the last three intervention topics. A significant interaction between condition and topic was detected, F (3.717, 486.984) = 2.743, $p = .032$, Partial $\eta^2 = 0.021$, $\varepsilon = 0.929$. Follow-up tests indicated significant simple effects for condition in all three topics. Again, acceptance of statistical significance was set at $p < .017$.

At Topic 2, the percentage of Shared evidence was significantly lower in the Control condition compared to the DD condition ($M = -0.08$, $SE = 0.03$, $p = .016$) and PD condition ($M = -0.27$, $SE = 0.02$, $p < .0005$). In addition, the percentage of Shared evidence was significantly higher in the PD condition compared to the DD condition ($M = 0.19$, $SE = 0.03$, $p < .0005$).

At Topic 3, the percentage of Shared evidence was significantly lower in the Control condition compared to the PD condition ($M = -0.19$, $SE = 0.02$, $p < .0005$), but not to the DD condition ($M = -0.04$, $SE = 0.02$, $p = .122$). In addition, the percentage of Shared evidence was significantly higher in the PD condition compared to the DD condition ($M = 0.16$, $SE = 0.02$, $p < .0005$).

At Topic 4, the percentage of Shared evidence was significantly lower in the Control condition compared to the PD condition ($M = -0.20$, $SE = 0.02$, $p < .0005$), but not to the DD condition ($M = -0.02$, $SE = 0.02$, $p = .313$). In addition, the percentage of Shared evidence was significantly higher in the PD condition compared to the DD condition ($M = 0.19$, $SE = 0.02$, $p < .0005$).

3.2. Pre- and post-assessment essays

A one-way ANOVA with condition as a predictor variable was carried out for pre-assessment essays to establish initial equivalence

Table 6
Percentage of students who ever included a type of integration strategy by condition and topic.

Category of Integration	Pre-assessment		Topic 1		Topic 2		Topic 3		Topic 4		Post-assessment	
	PD (N = 100)	DD (N = 68)	PD (N = 100)	DD (N = 68)	PD (N = 100)	DD (N = 68)	PD (N = 100)	DD (N = 68)	PD (N = 100)	DD (N = 68)	PD (N = 100)	DD (N = 68)
Dismissal	5.0	2.9	8.0	4.4	4.0	11.8	7.0	19.1	5.0	17.6	6.0	19.1
Refutation	2.0	4.4	2.0	2.9	4.0	5.9	37.0	13.2	25.0	7.4	19.0	13.2
Weighing/ Synthesis	0	0	0	0	0	0	5.0	5.9	5.0	2.9	9.0	11.8

Table 7
Mean percentage (and standard deviations) of functional evidence-based units by condition and topic.

Condition	Source of evidence	Pre-assessment	Topic 1	Topic 2	Topic 3	Topic 4	Post-assessment
PD condition	Personal evidence	11.36 (10.06)	12.39 (15.07)	13.90 (15.36)	10.93 (11.24)	11.78 (11.68)	12.86 (14.23)
	Shared evidence	9.96 (9.31)	0	31.80 (23.80)	27.61 (18.18)	29.22 (15.06)	25.96 (17.11)
DD condition	Personal evidence	13.35 (13.08)	12.11 (14.71)	11.65 (12.90)	13.86 (13.63)	12.95 (14.09)	12.10 (12.17)
	Shared evidence	12.95 (12.51)	0	12.73 (13.15)	11.86 (12.11)	10.67 (9.37)	10.90 (9.58)
Control condition	Personal evidence	13.45 (14.67)	12.93 (10.49)	12.14 (10.41)	15.21 (13.24)	14.19 (12.89)	13.90 (11.91)
	Shared evidence	8.33 (9.10)	0	5.21 (8.77)	8.37 (10.52)	8.77 (9.52)	8.33 (9.10)

across conditions in the mean number of idea units, the mean percentage of position-congruent and position-incongruent arguments, the percentage of students who ever employed each integration strategy, and the mean percentage of evidence-based units. None of these analyses revealed significant condition difference, suggesting that students from the three conditions demonstrated comparable performance in essay writing at pre-assessment.

Next, a one-way analysis of covariance (ANCOVA) with condition as a predictor variable and pre-assessment as a covariate was applied to compare performance across conditions at post-assessment. For position-congruent arguments, there was a statistically significant difference between conditions, F (2, 261) = 10.159, $p < .0005$, Partial $\eta^2 = 0.072$. Post hoc analysis was performed with a Bonferroni adjustment. The mean percentage of position-congruent arguments was significantly lower in the PD condition compared to the Control condition, with a mean difference of 0.07 (95 % CI, 0.032–0.116), $p = .001$. The mean percentage of position-congruent arguments was also significantly lower in the DD condition compared to the Control condition, with a mean difference of 0.098 (95 % CI, 0.051–0.144), $p < .0005$. No significant difference was found between the PD and DD conditions ($p = .309$).

A significant difference across conditions was also detected for position-incongruent arguments at post-assessment, F (2, 261) = 13.595, $p < .0005$, Partial $\eta^2 = 0.094$. Post hoc analysis was performed with a Bonferroni adjustment. The mean percentage of position-incongruent arguments was significantly higher in the PD condition compared to the Control condition, with a mean difference of 0.039 (95 % CI, 0.020–0.058), $p < .0005$. The mean percentage of position-incongruent arguments was also significantly higher in the DD condition compared to the Control condition, with a mean difference of 0.051 (95 % CI, 0.030–0.072), $p < .0005$. No significant difference was found between the PD and DD conditions ($p = .282$). Students never used *Dialogue reference* statements to introduce a position-incongruent argument at post-assessment, possibly because they did not engage in peer dialogues prior to writing.

In terms of the integration of position-incongruent arguments, Chi-square analysis at the individual level indicated that at post-assessment, significantly more DD than PD students used *Dismissal* at least once ($p = .012$), while the percentages of students who used *Refutation* ($p = .401$) or *Weighing/Synthesis* ($p = .608$) were comparable. Further analysis revealed that 90 % of *Refutation* was *Direct Refutation* in the PD condition, compared to only 15 % in the DD condition.

For evidence-based arguments, there was no significant condition difference in the mean percentage of Personal evidence, F (2, 261) = 0.456, $p = .635$, Partial $\eta^2 = 0.003$. However, a significant condition difference was found in the mean percentage of Shared evidence, F (2, 261) = 51.762, $p < .0005$, Partial $\eta^2 = 0.284$. Post hoc analysis was performed with a Bonferroni adjustment. The mean percentage of Shared evidence was significantly higher in the PD condition compared to the Control condition, with a mean difference of 0.176 (95 % CI, 0.140–0.212), $p < .0005$. The mean percentage of Shared evidence was also significantly higher in the PD condition

Table 8
Essay from PD condition student (Position: Support Yao's parents).

Idea unit	Essay content	Argumentative function and evidence use	Strategy to introduce and integrate position-incongruent arguments
1	Dear Yao, I heard that you had an argument with your parents and I want to say something.	Position-congruent	
2	I think your parents are right because you are too young.	Position-congruent	
3	My partner and I discussed how you have the right to use cell phones and laptops at home.	Position-incongruent	Dialogue reference; Advanced Refutation
4	But you don't have that right because you are too young to know what's right and what's not.	Position-congruent	
5	You are already getting problems with your eyes because you use your phones too much.	Position-congruent	
6	You can even hurt your neck and body if you do this longer.	Position-congruent; Shared evidence	
7	You spend less time with friends because you are with your phone all the time.	Position-congruent	
8	The evidence also said using too much social media might cause loneliness and depression and that's not what you want.	Position-congruent; Shared evidence	

compared to the DD condition, with a mean difference of 0.150 (95 % CI, 0.110–0.190), $p < .0005$. No significant difference was found between the DD and Control conditions ($p = .205$).

3.3. Qualitative analysis: case study of essays from Topic 3

Here we present a qualitative analysis comparing essays from the three conditions to shed further light on the above quantitative findings. We selected three representative essays from Topic 3, one from each condition, as shown in Tables 8 to 10. For all three essays, a considerable proportion of idea units were coded as position-congruent arguments, consistent with prior research demonstrating a predominance of claims consistent with one's position in student essays. Differences were observed, however, in students' use of position-incongruent arguments: in the PD condition essay (Table 8), the author referred to the preceding dialogue when introducing a position-incongruent argument in Unit 3 and countered it with a *Direct Refutation* in Unit 4; in the DD condition essay (Table 9), although the author also cited position-incongruent arguments in Units 7 and 8, she dismissed both of them; in the Control condition essay (Table 10), however, the author did not include any position-incongruent arguments. Differences in evidence use were also observed in the three essays: the PD condition essay drew on a piece of Shared evidence; the DD condition essay drew on a piece of Personal evidence and the Control condition essay contained no reference to either Personal or Shared evidence.

4. Discussion

The present study employed a pre- and post-test quasi-experimental design to investigate the impact of discourse goals on argumentative writing in elementary school students. Our analysis corroborated prior studies in showing that extended practice of direct peer-to-peer dialogues was superior to traditional teacher-led discussion in supporting the development of argument skills in argumentative writing, with gains manifested both in intervention topics and in a non-discourse transfer topic (Iordanou & Kuhn, 2020; Kuhn & Crowell, 2011; Kuhn et al., 2016; Rapanta, 2021).

The present study made at least two contributions to existing literature: first, it focused on elementary school students, a population generally ignored in prior studies that sought to improve students' argumentation skills by manipulating discourse goals; second, it showed that elementary school students might derive more benefits from participating in argumentative dialogues with the goal to persuade than to reach consensus through deliberation, a finding contradicting prior studies that focused on older students at the secondary or college levels (Asterhan & Babichenko, 2015; Felton et al., 2009; Felton et al., 2015a, 2015b; Villarroel et al., 2016). Therefore, we argue that teachers should remain sensitive to students' age level when selecting discourse goals for argumentative peer dialogues.

4.1. Differences in introducing and integrating position-incongruent arguments

We first compare performances between the two intervention conditions and the control condition. In the last two intervention topics, the percentage of position-incongruent arguments was significantly higher in the PD or DD conditions than in the Control condition, suggesting that students who participated in argumentative dialogues were more likely to insert mention of arguments in favor of an opposing position. Although all the students received the same set of evidence in support of conflicting positions when writing essays, the mere exposure to conflicting evidence, without engagement in direct peer-to-peer argumentative dialogues, was not enough to prompt Control students to generate position-incongruent arguments (Iordanou & Kuhn, 2020). In addition, the finding that significant condition difference was observed in later topics but not in earlier topics possibly suggested that dialogues had a cumulative effect on students as they continued to engage in them, although we cannot completely rule out topic effects when explaining this finding. For the non-discourse topic at post-assessment, PD and DD students again outperformed Control students in including position-inconsistent arguments, suggesting that intervention students were able to abstract and transfer their newly developed skills to acknowledge and address position-inconsistent arguments to a novel topic (Nussbaum & Asterhan, 2016; Reznitskaya et al., 2001, 2012).

Next, we contrast the performance of the two intervention conditions in integrating position-incongruent arguments. In the last two intervention topics, a significantly higher percentage of PD students ever employed the *refutation* strategy, while a significantly higher percentage of DD students ever employed the *dismissal* strategy. This condition difference was carried into the post-assessment task, as significantly more DD than PD students ever used the *Dismissal* strategy. In addition, across all the intervention topics and post-assessment topic, PD students were more likely to employ the stronger *Refutation* strategy to directly weaken the conclusion or premise of a position-inconsistent argument, while DD students were more likely to employ the weaker *Refutation* strategy that proposed an alternative argument and thus left the original claim intact.

Although prior literature suggests *refutation* reflects an adversarial form of argumentation (Andriessen et al., 2003) and does not constitute the most integrative strategy (Casado-Ledesma et al., 2021; Mateos et al., 2018; Nussbaum & Schraw, 2007), we consider *refutation* a more successful strategy than *dismissal*, as the former was more cognitively demanding and represented a more advanced understanding of the argument structure (Toulmin, 1958). Moreover, PD students demonstrated an enhanced argumentative competence by directly weakening the conclusion or premise of a position-inconsistent claim, while DD students tended to leave the force of the original claim unaddressed.

We speculate that during argumentative dialogues, students aiming to persuade an opponent experienced greater cognitive conflict and had increased opportunities to critically question and evaluate each other's arguments, either by directly weakening an opposing-side argument or reinforcing an position-congruent argument attacked by an opponent, both of which promoted the formation and

Table 9
Essay from DD condition student (Position: Support Yao).

Idea unit	Essay content	Argumentative function and evidence use	Strategy to introduce and integrate position-incongruent arguments
1	Hello Yao, I think you are free to use your cellphones or laptops at home.	Position-congruent	
2	This topic is relevant to all of us and we all face it at home.	Position-congruent	
3	I believe you are old enough to plan your activities at home.	Position-congruent	
4	You can control what you do with your phone.	Position-congruent	
5	You can learn a lot from your phone	Position-congruent	
6	But your parents are right to be worried about your eyesight.	Position-incongruent;	No reference; Dismissal
7	I read somewhere that you could get addicted to your cellphones easily.	Position-incongruent; Personal evidence	General-other reference; Dismissal
8	If you don't know what to do, talk to your parents.	Other	

Table 10
Essay from Control condition student (Position: Support Yao's parents).

Idea unit	Essay content	Argumentative function and evidence use
1	Dear Yao, I think your parents should decide how you use your laptops and cellphones at home.	Position-congruent
2	I am saying so because they are your parents.	Position-congruent
3	Your grades at school will get better if you spend more time studying than using cellphones for fun.	Position-congruent
4	Your eyesight will also get better if you spend more time playing outside with friends.	Position-congruent
5	You can also spend more time with your parents if you don't use your cellphones or watch TV too much.	Position-congruent
6	I hope you know what I mean and I really think you should listen to your parents.	Position-congruent

consolidation of the argument-counterargument-rebuttal exchange structure (Leitão, 2000). Students aiming to deliberate, however, in their pursuit of consensus or solutions rather than differences of opinion, might have engaged less often in the direct exchange of arguments and counterarguments, as doing so would run contrary to the goal of arriving at a mutually agreed-upon opinion or collaboratively exploring solutions to a problem. Instead, DD students might have engaged in more talk involving quick consensus building (Lind et al., 2023; Thiebach et al., 2016; Weinberger & Fischer, 2006), which precluded in-depth exploration of opposing arguments and evidence.

In fact, the finding related to how students introduced a position-incongruent argument could provide some support for this interpretation. In the last two intervention topics, PD students were more likely to attribute a position-incongruent argument to the preceding dialogues. We speculate that PD students likely focused more on exploring differences of opinion during dialogues, the process of which made the origin of a position-incongruent argument salient and important to them. A critical next step is to analyze students' argumentative dialogues (Shi, 2020b) to reveal whether PD and DD students engaged in different patterns of dialogue, which would provide an explanatory account for the differences observed in essay writing as reported here.

Lastly, we should keep in mind that all the students showed minimal use of the most integrative strategies – *weighing* and *synthesis* – across topics. This finding was unsurprising, as prior studies on secondary school (Casado-Ledesma, 2021) and college students (Mateos et al., 2018) showed that when teaching students to integrate conflicting information sources in argumentative writing, teachers needed to add explicit instructional practices to dialogic, collaborative activities to enhance students' use of these advanced integrative strategies. Therefore, a possible next step is to examine whether elementary school students would derive more benefits from a combination of argumentative dialogues and direct instruction on the use of the more advanced *Weighing* and *Synthesis* strategies.

4.2. Differences in making evidence-based arguments

Our analyses also revealed enhanced performance of PD students in making evidence-based claims. While no condition difference was observed in the extent to which students drew on their personal knowledge to justify claims, significant condition difference was observed in students' use of evidence shared by the teacher. In essays at Topics 2, 3, 4 and the post-assessment essays, around 30 % of claims made by PD students were based on evidence shared by the teacher, far exceeding that by DD and Control students. We speculate that during argumentative dialogues, PD students were more motivated to draw on provided evidence to support or weaken claims, possibly because doing so might help them strengthen their own side or reduce the force of the opposing side, both of which would contribute to the goal of winning the debate (Walton, 1989). DD students, on the other hand, were less motivated to draw on provided evidence to support or weaken claims, as doing so might diverge from the goal of exploring solutions to satisfy both parties. It is noticeable that PD students transferred their ability to coordinate claims with evidence to the post-assessment task. These findings were particularly encouraging given the accumulation of research indicating a strong tendency across age groups to substitute mechanism explanation for empirical evidence when supporting a causal claim (Ahn et al., 1995; Brem & Rips, 2000; Kuhn, 1991;

Kuhn & Pearsall, 1998, 2000; Ross et al., 1975). Our study suggested that engaging in argumentative dialogues with the goal to win prompt elementary school students to overcome this tendency to a greater extent.

5. Conclusion

We would like to end by reiterating one of the most prominent findings in the present study. Although deliberation dialogues hold the potential to decrease defense motivation and increase balanced consideration of opposing-side arguments, for elementary school students, persuasion dialogues might lead to more extended and successful exchange of arguments and counterarguments, as well as deeper exploration and greater use of evidence to justify claims. Therefore, rather than consistently favoring deliberation over persuasion as recommended in studies with older populations (Asterhan & Schwarz, 2016; Felton et al., 2009; Felton et al., 2015a, 2015b; Nussbaum & Edwards, 2011), elementary school teachers should be more prudent in choosing the most appropriate discourse goal. Arguing to persuade – a goal children are more familiar with – might be a desirable and productive starting point and over time, teachers could introduce the goal of deliberation or alternate between the two goals so that students could derive benefits from both.

Indeed, aside from holding dialogues with disagreeing peers, students could also discuss with agreeing peers (Iordanou & Kuhn, 2020; Kuhn et al., 1997, 2019), as conversing with opposing-side and same-side peers might lead to different discourse patterns and thus diversified gains in argument skills. In addition, by focusing on argumentative dialogues, the present study did not contrast the dialogue-based approach with other instructional methods reported effective in improving students' argumentative writing (Graham & Perin, 2007), such as engaging students in collaborative discussions and practices (Wissinger & De La Paz, 2016), focusing on students' self-regulated strategy development (Ferretti & Graham, 2019), providing cognitive or metacognitive instruction (De La Paz & Felton, 2010; Maier & Richter, 2014), employing graphic organizers or diagrams (Barzilai & Ka'adan, 2017; Nussbaum, 2008), or combining two or more of these instructional activities (Granado-Peinado et al., 2019; Mateos et al., 2018). Future studies could investigate which (or what combination) of these approaches is most effective in improving students' argumentative writing skills.

Funding information

This research was supported by the National Social Science Foundation for Education of China (CEA200257) awarded to the author.

Ethics and integrity statements

Data for this research is available upon request.

CRediT authorship contribution statement

Yuchen Shi: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

References

Anderson, R. C. (1984). A schema-theoretic view of basic processes in reading comprehension. In P. D. Pearson, P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 255–291). Longman, Inc.

Andriessen, J., Baker, M., & Suthers, D. (2003). *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments*. Kluwer Academic.

Applebee, A. N., Langer, J. A., Mullis, I. V. S., & Jenkins, L. (1990). *The writing report card, 1984–88: Findings from the nation's report card*. National Assessment of Educational Progress.

Applebee, A. N. (1996). *Curriculum as conversation: Transforming traditions of teaching and learning*. The University of Chicago Press.

Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal, 40*(3), 685–730.

Asterhan, C. S. C., & Babichenko, M. (2015). The social dimension of learning through argumentation: Effects of human presence and discourse style. *Journal of Educational Psychology, 107*(3), 740–755.

Asterhan, C. S. C., & Schwarz, B. B. (2007). The effects of monological and dialogical argumentation on concept learning in evolutionary theory. *Journal of Educational Psychology, 99*(3), 626–639.

Asterhan, C., & Schwarz, B. (2016). Argumentation for learning: Well-trodden paths and unexplored territories. *Educational Psychologist, 51*, 164–187.

Baker, M. J. (2003). Computer-mediated interactions for the co-elaboration of scientific notions. In J. Andriessen, M. Baker, & D. Suthers (Eds.), *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments*. Kluwer Academic.

Bakhtin, M. M. (1981). *The dialogic imagination: Four essays by M. M. Bakhtin*. University of Texas Press.

Baron, J. (1988). *Thinking and deciding*. Cambridge University Press.

Barzilai, S., & Ka'adan, I. (2017). Learning to integrate divergent information sources: The interplay of epistemic cognition and epistemic metacognition. *Metacognition and Learning, 12*(2), 193–232.

Bell, P., & Linn, M. C. (2000). Scientific arguments as learning artifacts: Designing for learning from the web with KIE. *International Journal of Science Education, 22*, 797–817.

Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education, 94*(5), 765–793.

Berland, L. K., & Reiser, B. J. (2011). Classroom communities' adaptations of the practice of scientific argumentation. *Science Education, 95*(2), 191–216.

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

Brem, S. K., & Rips, L. J. (2000). Explanation and evidence in informal argument. *Cognitive Science, 24*, 573–604.

Clark, D. B., & Sampson, V. (2007). Personally-seeded discussions to scaffold online argumentation. *International Journal of Science Education, 29*(3), 253–277.

Coirier, P. (1996). Composing argumentative texts: Cognitive and/or textual complexity. In G. Rijlaarsdam, H. van der Bergh, & & M. Couzijn (Eds.), *Current trends in research on writing: Theories, models, and methodology* (pp. 317–338). Amsterdam University Press.
Casado-Ledesma, L., Cuevas, I., Van den Bergh, H., Rijlaarsdam, G., Mateos, M., Granado-Peinado, M., et al. (2021). Teaching argumentative synthesis writing through deliberative dialogues: Instructional practices in secondary education. *Instructional Science, 49*(4), 515–559.
Crowell, A., & Kuhn, D. (2014). Developing dialogic argumentation skills: A three-year intervention study. *Journal of Cognition and Development, 31*, 456–496.
De La Paz, S., & Felton, M. K. (2010). Reading and writing from multiple source documents in history: Effects of strategy instruction with low to average high school writers. *Contemporary Educational Psychology, 35*(3), 174–192.
De La Paz, S., Ferretti, R., Wissinger, D., Yee, L., & MacArthur, C. (2012). Adolescents' disciplinary use of evidence, argumentative strategies, and organizational structure in writing about historical controversies. *Written Communication, 29*(4), 412–454.
Eemeren, F., Grootendorst, R., & Kruiger, T. (1987). *Handbook of argumentation theory: A critical survey of classical backgrounds and modern studies*. De Gruyter Mouton.
Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education, 88*(6), 915–933.
Feilke, H. (1996). From syntactical to textual strategies of argumentation. *Argumentation, 10*, 197–212.
Felton, M., Crowell, A., & Liu, T. (2015a). Arguing to agree: Mitigating my-side bias through consensus-seeking dialogue. *Written Communication, 32*, 317–331.
Felton, M., Garcia-Mila, M., Villarroel, C., & Gilabert, S. (2015b). Arguing collaboratively: Argumentative discourse types and their potential for knowledge building. *British Journal of Educational Psychology, 85*, 372–386.
Ferretti, R. P., & Graham, S. (2019). Argumentative writing: Theory, assessment, and instruction. *Reading and Writing: An Interdisciplinary Journal, 32*(6), 1345–1357.
Ferretti, R. P., MacArthur, C. A., & Dowdy, N. S. (2000). The effects of an elaborated goal on the persuasive writing of students with learning disabilities and their normally achieving peers. *Journal of Educational Psychology, 92*(4), 694–702.
Golanics, J. D., & Nussbaum, E. M. (2008). Enhancing online collaborative argumentation through question elaboration and goal instructions. *Journal of Computer Assisted Learning, 24*(3), 167–180.
Golder, C. (1992). Production of elaborated argumentative discourse: The role of cooperativeness. *European Journal of Psychology of Education, 7*, 51–59.
Golder, C., & Coirier, P. (1994). Argumentative text writing: Developmental trends. *Discourse Processes, 18*, 187–210.
Golder, C., & Coirier, P. (1996). The production and recognition of typological markers. *Argumentation, 10*, 271–282.
Graff, G. (2003). *Clueless in academe: How schooling obscures the life of the mind*. Yale University Press.
Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*(3), 445–476.
Granado-Peinado, M., Mateos, M., Martín, E., & Cuevas, I. (2019). Teaching to write collaborative argumentative syntheses in higher education. *Reading and Writing, 32*(8), 2037–2058.
Hemberger, L., Kuhn, D., Matos, F., & Shi, Y. (2017). A dialogic path to evidence-based argumentive writing. *Journal of the Learning Sciences, 26*(4), 575–607.
Iordanou, K., & Constantinou, C. (2015). Supporting use of evidence in argumentation through practice in argumentation and reflection in the context of SOCRATES learning environment. *Science Education, 99*(2), 282–311.
Iordanou, K., & Kuhn, D. (2020). Contemplating the opposition: Does a personal touch matter? *Discourse Processes, 57*(4), 343–359.
Kuhn, D. (1991). *The skills of argument*. Cambridge University Press.
Kuhn, D. (2022). Metacognition matters in many ways. *Educational Psychologist, 57*(2), 73–86.
Kuhn, D., & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents' thinking. *Psychological Science, 22*, 545–552.
Kuhn, D., Floyd, D., Yaksick, P., Halpern, M., & Ricks, W. (2019). How does discourse among like-minded individuals affect their thinking about a complex issue? *Thinking & Reasoning, 25*(3), 365–382.
Kuhn, D., Hemberger, L., & Khait, V. (2016). Tracing the development of argumentive writing in a discourse-rich context. *Written Communication, 33*, 92–121.
Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentative reasoning. *Cognition and Instruction, 15*(3), 287–315.
Leitão, S. (2000). The potential of argument in knowledge building. *Human Development, 43*, 323–360.
Leitão, S. (2003). Evaluating and selecting counterarguments: Studies of children's rhetorical awareness. *Written Communication, 20*(3), 269–306.
Lind, A., Andersson-Bakken, E., & Sandvik, M. (2023). Patterns of peer talk in consensus-oriented classrooms: Deliberative argumentation or rush toward consensus. *Learning, Culture, and Social Interaction, 40*, Article 100703.
Maier, J., & Richter, T. (2014). Fostering multiple text comprehension: How metacognitive strategies and motivation moderate the text-belief consistency effect. *Metacognition and Learning, 9*(1), 51–74.
Mateos, M., Martín, E., Cuevas, I., Villalón, R., Martínez, I., & González-Lamas, J. (2018). Improving written argumentative synthesis by teaching the integration of conflicting information from multiple sources. *Cognition and Instruction, 36*(2), 119–138.
Mead, G. H. (1972). *Mind, self, and society. from the standpoint of a social behaviorist* (18th ed.). The University of Chicago Press (Original work published 1934).
Mehan, H. (1979). *Learning lessons: Social organization in the classroom*. Harvard University Press.
Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34*, 57–111.
Moshman, D. (2011). *Adolescent rationality and development: Cognition, morality, and identity* (3rd ed.). Psychology Press.
Newell, G. E., Beach, R., Smith, J., & VanDerHeide, J. (2011). Teaching and learning argumentative reading and writing: A review of research. *Reading Research Quarterly, 46*(3), 273–304.
Nussbaum, E. M. (2008). Using argumentation vee diagrams (AVDs) for promoting argument-counterargument integration in reflective writing. *Journal of Educational Psychology, 100*(3), 549–565.
Nussbaum, E. M., & Asterhan, C. (2016). The psychology of far transfer from classroom argumentation. In F. Paglieri (Ed.), *The psychology of argumentation*. College publications.
Nussbaum, E. M., & Edwards, O. V. (2011). Critical questions and argument stratagems: A framework for enhancing and analyzing students' reasoning practices. *Journal of the Learning Sciences, 20*(3), 443–488.
Nussbaum, E. M., & Kardash, C. M. (2005). The effects of goal instructions and text on the generation of counterarguments during writing. *Journal of Educational Psychology, 97*(2), 157–169.
Nussbaum, E. M., & Schraw, G. (2007). Promoting argument-counterargument integration in students' writing. *Journal of Experimental Education, 76*(1), 59–92.
Osborne, J. F., Henderson, J. B., MacPherson, A., Szu, E., Wild, A., & Yao, S. Y. (2016). The development and validation of a learning progression for argumentation in science. *Journal of Research in Science Teaching, 53*(6), 821–846.
Perkins, D. N. (1985). Postprimary education has little impact on informal reasoning. *Journal of Educational Psychology, 77*, 562–571.
Perkins, D. N., Farady, M., & Bushey, B (1991). Everyday reasoning and the roots of intelligence. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), *Informal reasoning and education*. Erlbaum.
Pollock, J. L. (1987). Defeasible reasoning. *Cognitive Science, 11*, 481–518.
Rapanta, C. (2021). Can teachers implement a student-centered dialogical argumentation method across the curriculum? *Teaching and Teacher Education, 105*, Article 103404.
Rapanta, C., & Felton, M. K. (2022). Learning to argue through dialogue: A review of instructional approaches. *Educational Psychology Review, 34*, 477–509.
Reznitskaya, A., Anderson, R. C., McNurlen, B., Nguyen-Jahiel, K., Archodidou, A., & Kim, S. (2001). Influence of oral discussion on written argument. *Discourse Processes, 32*(2/3), 155–175.
Reznitskaya, A., Glina, M., Carolan, B., Michaud, O., Rogers, J., & Sequeira, L. (2012). Examining transfer effects from dialogic discussions to new tasks and contexts. *Contemporary Educational Psychology, 37*(4), 288–306.
Ryu, S., & Sandoval, W. A. (2012). Improvements to elementary children's epistemic understanding from sustained argumentation. *Science Education, 96*(3), 488–526.
Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction, 23*(1), 23–55.
Shi, Y. (2019). Enhancing evidence-based argumentation in a Mainland China middle school. *Contemporary Educational Psychology, 59*, Article 101809.
Shi, Y. (2020a). Constructed dialogs reveal skill development in argumentive writing. *Reading and Writing, 33*(9), 2311–2335.

Shi, Y. (2020b). Talk about evidence during argumentation. *Discourse Processes, 57*(9), 770–792.
Shi, Y., Shen, X., Wang, T., Cheng, L., & Wang, A. (2021). Dialogic teaching of controversial public issues in a Chinese middle school. *Learning, Culture and Social Interaction, 30*, Article 100533.
Shi, Y., Zhang, Z., Cao, S., & Liu, Q. (2024). Dialogic teaching of controversial issues: Discursive moves to enact two-sided discussions. *Language and Education, 38*(2), 303–319.
Sinclair, J. M. H., & Coulthard, M. (1975). *Towards an analysis of discourse: The English used by pupils and teachers*. Oxford University Press.
Stanovich, K. E., & West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Thinking & Reasoning, 13*(3), 225–247.
Stein, N. L., & Bernas, R. S. (1999). The early emergence of argumentative knowledge and skill. G. Rijlaarsdam & E. Espéret (Series Eds.) and J. Andriessen & P. Coirier (Vol. Eds.). *Studies in writing: Vol. 5. foundations of argumentative text processing* (pp. 97–116). Amsterdam University Press.
Stein, N. L., & Miller, C. A. (1991). I win—You lose: The development of argumentative thinking. In J. F. Voss, D. N. Perkins, & Judith W. Segal (Eds.), *Informal reasoning and education* (pp. 265–290). Erlbaum.
Thiebach, M., Mayweg-Paus, E., & Jucks, R. (2016). Better to agree or disagree? The role of critical questioning and elaboration in argumentative discourse. *Zeitschrift für Pädagogische Psychologie /German Journal of Educational Psychology, 30*(2–3), 133–149.
Toulmin, S. (1958). *The uses of argument*. Cambridge University Press.
VanDerHeide, J., & Newell, G. E. (2013). Instructional chains as a method for examining the teaching and learning of argumentative writing in classrooms. *Written Communication, 30*(3), 300–329.
Villarroel, C., Felton, M., & Garcia-Mila, M. (2016). Arguing against confirmation bias: The effect of argumentative discourse goals on the use of disconfirming evidence in written argument. *International Journal of Educational Research, 79*, 167–179.
Vygotsky, L. (1962). *Thought and language*. MIT Press.
Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
Walton, D. (1989). Dialogue theory for critical thinking. *Argumentation, 3*, 169–184.
Walton, D., & Krabbe, E. (1995). *Commitment in dialogue*. State University of New York Press.
Walton, D., & Zhang, N. (2013). The epistemology of scientific evidence. *Artificial Intelligence & Law, 21*(2), 173–219.
Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46*(1), 71–95.
Wertsch, J. V. (1991). *Voices of the mind*. Harvard University Press.
Wissinger, D. R., & De La Paz, S (2016). Effects of critical discussions on middle school students' written historical arguments. *Journal of Educational Psychology, 108*(1), 43–59.
Wolfe, C. R. (2011). Argumentation across the curriculum. *Written Communication, 28*(2), 193–219.
Wolfe, C. R. (2012). Individual difference in the "myside bias" in reasoning and written argumentation. *Written Communication, 29*(4), 477–501.
Wolfe, C. R., Britt, M. A., & Butler, J. A. (2009). Argumentation schema and the Myside bias in written argumentation. *Written Communication, 26*(2), 183–209.
Zavala, J., & Kuhn, D. (2017). Solitary discourse is a productive activity. *Psychological Science, 28*(5), 578–586.
Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching, 39*(1), 35–62.