

Central Lancashire Online Knowledge (CLoK)

Title	A Dialogic Approach to Enhancing Argumentative Writing From Multiple Texts
Type	Article
URL	https://clock.uclan.ac.uk/54258/
DOI	https://doi.org/10.1177/20965311241310884
Date	2025
Citation	Shi, Yuchen and Iordanou, Kalypso (2025) A Dialogic Approach to Enhancing Argumentative Writing From Multiple Texts. ECNU Review of Education. ISSN 2096-5311
Creators	Shi, Yuchen and Iordanou, Kalypso

It is advisable to refer to the publisher's version if you intend to cite from the work.
<https://doi.org/10.1177/20965311241310884>

For information about Research at UCLan please go to <http://www.uclan.ac.uk/research/>

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the <http://clock.uclan.ac.uk/policies/>

A Dialogic Approach to Enhancing Argumentative Writing From Multiple Texts

ECNU Review of Education

1–27

© The Author(s) 2025

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/20965311241310884

journals.sagepub.com/home/roe**Yuchen Shi** (石雨晨) 

Institute of Curriculum and Instruction, East China Normal University

Kalypso Iordanou

University of Central Lancashire Cyprus

Abstract

Purpose: The study sought to examine the effectiveness of a dialogue-based argument intervention in enhancing Chinese middle school students' integration of conflicting information from multiple texts in argumentative writing.

Design/Approach/Methods: The study followed a quasi-experimental design with pre-assessment and post-assessment, comparing seventh-grade intervention and non-participating control students' individual post-assessment writing performance on a non-discourse topic involving genetically modified foods.

Findings: Intervention students outperformed control students in integrating textual evidence inconsistent with one's position. Specifically, intervention students were more successful in integrating position-inconsistent information with their prior knowledge or integrating multiple pieces of position-inconsistent information from one text or across multiple texts. Intervention students were also more successful in integrating two pieces of conflicting information. When judging text trustworthiness, intervention students trusted a primary source to a greater extent and showed greater gains in taking into consideration the epistemological aspect, as well as one's own or a text's position on the issue.

Corresponding author:

Yuchen Shi, Institute of Curriculum and Instruction, Faculty of Education, East China Normal University, 3663 North Zhongshan Road, Shanghai 200062, China.

Email: ycshi@kcx.ecnu.edu.cn



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Originality/Value: The present study demonstrated the effectiveness of the dialogue-based argument curriculum in promoting Chinese middle school students' ability to write integrated essays from multiple texts.

Keywords

Argumentative dialogue, argumentative writing, Chinese students, controversial issues, evidence, multiple texts

Date received: 10 March 2023; revised: 6 May 2024; accepted: 29 June 2024

The digital revolution of the past few decades brings with it a host of demands necessary to live and learn well in the 21st century, including the need to navigate the unprecedented quantity and complexity of diverse information sources, in order to learn about controversial topics, “for which there is a scientific knowledge base but about which there is controversy in the public domain, making it highly likely that the lay public will encounter conflicting points of view on these topics” (Goldman & Scardamalia, 2013, p. 256). However, students across educational levels often engage in ineffective strategies when processing divergent sources (Britt & Aglinskias, 2002; Wineburg, 1991), constructing only a sufficient rather than the best possible representation of a controversial issue (Richter & Maier, 2017).

The present work aimed to examine the effectiveness of a dialogue-based argument intervention (Kuhn et al., 2016; Kuhn et al., 2024) in supporting Chinese middle school students' integration of evidence from conflicting sources in argumentative writing. Following the sociocultural approach (Vygotsky, 1978), the present research employed peer dialogues that focused on persuasive argumentation to enhance students' integration of conflicting information from multiple texts in writing on a controversial topic. In doing so, we extended the range of instructional activities employed in prior research that aimed to support students' integrated writing from multiple information sources (Barzilai et al., 2018). Another contribution of the present research was that we examined whether and how an argument-focused intervention would enhance students' judgment of text trustworthiness when dealing with multiple texts and the reasoning process they engaged in. Following prior research, we defined multiple texts as a collection of texts addressing the same controversial topic (e.g., Barzilai et al., 2018; Stahl et al., 1996), with some texts supporting one side and other texts an opposing side.

Integration of evidence from multiple texts in argumentative writing

Text-based argumentative writing requires individuals to construct and communicate a justified and balanced position on the basis of reasons and evidence presented across multiple information

sources (Kiili et al., 2020; Litman et al., 2017). Compared to working with multiple consistent sources or a single source that presents only one view on an issue, working with multiple conflicting sources allows individuals to build a deeper representation of a controversial issue (Bråten & Braasch, 2018; Kienhues et al., 2011). When students encounter multiple accounts of a controversial topic, however, the trustworthiness of each account is indeterminate and requires critical examination (Barzilai & Ka'adan, 2017; Kiili et al., 2020; Sinatra & Lombardi, 2020). Therefore, their decision regarding which account is credible enough to be included in discourse with others or in essay writing is critical.

Apart from judging the trustworthiness of multiple information sources, successful argumentative writing also requires a writer to include and elaborate on trustworthy textual information both consistent and inconsistent with one's position (Hagen et al., 2014). However, one-sided writing that fails to take into account competing arguments and the associated evidence has been prevalent among students across age groups (Kiili et al., 2020; Nussbaum & Schraw, 2007; O'Keefe, 1999; Reznitskaya et al., 2009; Wolfe et al., 2009).

To help students overcome this challenge, experimental studies that did and did not involve instructional interventions were carried out (Barzilai et al., 2018). For studies that did not involve instructional activities, one prominent line of research sought to manipulate task goals to improve students' text-based writing, such as randomly assigning students to write arguments, narratives, or summaries, following their reading of multiple texts (Bråten & Strømsø, 2009; Gil et al., 2010a, 2010b; Le Bigot & Rouet, 2007; Wiley & Voss, 1999). Studies that involved instructional activities included those that engaged students in collaborative discussions and practices (Wissinger & De La Paz, 2016), provided cognitive or metacognitive instruction related to integration strategies (De La Paz & Felton, 2010; Du & List, 2021; Maier & Richter, 2014), employed graphic organizers or diagrams (Barzilai & Ka'adan, 2017; Nussbaum, 2008), or combined two or more of these instructional activities (Granado-Peinado et al., 2019; Mateos et al., 2018). While some positive results were obtained from extant research, the production of written argumentative synthesis from multiple texts remains a challenging task, and task goal manipulations or instructional interventions do not always lead to enhanced performance (e.g., Barzilai & Ka'adan, 2017; Du & List, 2021; Gil et al., 2010a, 2010b).

Importantly, to successfully integrate conflicting information from multiple texts, a writer needs to attend to and resolve discrepancies that occurred within a text or across multiple texts, as well as between a source and one's prior knowledge, beliefs, or attitudes (Braasch & Kessler, 2021). However, students do not always recognize these discrepancies, and when they do, they tend to discount anomalous data in various ways to protect their pre-instructional theory (Chinn & Brewer, 1993) or preexisting beliefs (Klaczynski & Gordon, 1996). In fact, researchers have consistently demonstrated the existence of myside bias or confirmation bias (Nickerson, 1998) in written argumentation, manifested in a writer's consistent tendency to ignore or discount

discrepant information and thus fail to include any reference to other-side arguments or positions in writing (Kunda, 1990; O’Keefe, 1999; Wolfe & Britt, 2008; Wolfe et al., 2009). Although motivated reasoning (Kunda, 1990) that sought to protect one’s existing beliefs or viewpoints could have an evolutionary advantage when one tries to persuade others (Mercier & Sperber, 2011), to ignore or to exclude other-side information in one’s reasoning or writing is counterproductive, or even detrimental, when one attempts to form a coherent understanding or make critical decisions regarding controversial issues of personal and social significance (Shi et al., 2021; Shi et al., 2024a, 2024b).

To develop such skills, a potentially productive path would be to engage students in extended peer-to-peer dialogues focusing on adversarial argumentation, in which students are constantly called upon to undermine their opponents through the use of counterarguments, and to address challenges to one’s viewpoint through the use of rebuttals (Walton, 1989). According to the sociocultural framework (Vygotsky, 1978; Wertsch, 1979), these dialectic transactive moves (e.g., claim revision, articulation of opposing viewpoints, and argument–counterargument integration) would be gradually acquired and internalized to support students’ individual construction of integrated arguments that address contrary perspectives in subsequent writing activities (Rapanta & Felton, 2022). However, limited research has examined the effects of extended adversarial argumentation between opposing sides as an instructional approach to improve students’ integration of conflicting information from multiple texts in writing (Barzilai et al., 2018).

Judgment of text trustworthiness

As mentioned above, prior to selecting and incorporating textual information consistent or inconsistent with one’s position, a writer needs to first judge and decide on the trustworthiness of such information. Existing research indicated that characteristics of the readers themselves, such as prior beliefs, domain expertise, and epistemological understanding, might influence their ability and disposition to judge the trustworthiness of a text (Rouet et al., 1997; Rouet & Potocki, 2018; Strømsø et al., 2011). For example, while undergraduate students considered textbooks to be trustworthy sources, expert source users (e.g., graduate students and historians) tended to trust primary sources to a significantly greater extent (Bråten et al., 2009, 2011; Britt & Aglinskas, 2002; Rouet et al., 1997; Stahl et al., 1996; Wineburg, 1991).

In addition, Rouet et al. (1996) reported that students’ justificatory criteria varied across document types, with content characteristics invoked more often when evaluating textbooks and source characteristics (i.e., document type and author) invoked more often when evaluating primary documents. Bråten et al. (2011) further investigated how undergraduates judged the trustworthiness of different information sources bearing on a controversial topic. They found that students held information from textbooks and official documents (e.g., information from a university-based research

center and a government office) to be more trustworthy than information from newspapers and a commercial agent.

Specifically focusing on students' judgment of online sources, Kiili et al. (2022) examined upper secondary school students' credibility evaluation when reading texts on a controversial issue in a website-based environment. The authors reported that while students' credibility evaluations were quite accurate, their credibility justifications lacked sophistication. Hämäläinen et al. (2020) conducted intervention studies in which teachers delivered direct instruction, which included modeling, prompting, and discussing evaluation strategies, to improve sixth graders' performance in credibility evaluation of online sources. While the intervention led to improvement in students' skills to evaluate source features, no improvement in content-based evaluation of credibility was observed.

In spite of these studies, however, our understanding regarding the factors that affect middle school students' judgement of text trustworthiness, as well as how to encourage and train students to actively and effectively evaluate the trustworthiness of different information sources using relevant criteria (Bråten et al., 2011), is still quite limited. More work focusing, in particular, on the *reasoning behind* text trustworthiness judgments (List et al., 2017), and how this reasoning could be supported through interventions, is needed. Therefore, in addition to examining whether participation in the AWM intervention would facilitate students' integration of conflicting information from multiple texts in a post-assessment argumentative writing task, the present research also sought to investigate whether and how the intervention might affect students' judgment of text trustworthiness and the justificatory criteria they employed when multiple texts from various authentic sources on a controversial topic are provided.

The “Argue with me” curriculum

One prominent line of research that leveraged peer-to-peer adversarial dialogues to support individual argumentative writing (Hemberger et al., 2017; Kuhn & Crowell, 2011) was carried out by Kuhn and colleagues (Kuhn et al., 2008, 2016). Their dialogue-based argument curriculum, also called “Argue with me” (AWM), engaged primary and secondary school students in extended goal-based, dialogic activities and reflections with opposing-side peers on controversial issues of personal, societal, and scientific significance (Iordanou & Rapanta, 2021; Shi, 2019, 2024), before inviting students to write an individual argumentative essay on the issue. The AWM curriculum is based on Kuhn's theoretical framework (2000, 2001, 2022) on the development of argument skills, according to which argument skills is supported by strategic and metacognitive development, epistemological understanding, and intellectual dispositions to commit to these cognitively demanding practices.

One of the key theoretical underpinnings of the AWM curriculum was the view that thinking and reasoning are at heart dialogic (Billig, 1987; Cazden, 2001; Mead, 1934; Resnick et al., 2015;

Vygotsky, 1978). An argument, Gergen (2015) claims, depends for its meaning on how others respond. Others' reactions to my argument enrich and raise my confidence in its meaning. Graff (2003) argued that discourse provides students the "missing interlocutor" that often renders their expository writing devoid of purpose. An accumulating line of research following the AWM approach has demonstrated that gains in argument skills were first demonstrated in discourse with opposing-side peers (Crowell & Kuhn, 2014; Papathomas & Kuhn, 2017), and later in an interiorized form in individual essays, particularly in the critical respect of seeking to weaken opposing claims (Hemberger et al., 2017; Iordanou et al., 2019; Kuhn & Crowell, 2011; Rapanta, 2021; Shi, 2019, 2024). Although gains in the more challenging aspect of integrating arguments both consistent and inconsistent with one's position (Graham & Perin, 2007; Nussbaum & Schraw, 2007; Reznitskaya et al., 2001; Wolfe, 2011) in writing appeared much later and to a lesser extent, continued participation in the AWM curriculum supported the development of this crucial skill in writing (Matos, 2021; Shi, 2019, 2024). Indeed, compared to engaging in discourse with a same-side peer, arguing with an opposing-side peer has been demonstrated to lead to enhanced performance in addressing other-side claims in a post-discourse argumentative writing task (Iordanou & Kuhn, 2020).

Supporting strategic gains in argumentative discourse and writing are developments at the epistemological (Iordanou, 2017) and metacognitive levels (Kuhn et al., 2013; Shi, 2020a). In fact, development of argument skills involves acquiring better meta-level control of the application of effective argumentative strategies from individuals' repertory; at the same time, an individual is expected to acquire an increasing understanding that some strategies are more effective to achieve particular goals than others and gradually increase their use. Equally important to strong meta-level regulators for supporting the development of argument skills, according to Kuhn (2022), is the development of the disposition to engage in argumentation, which is closely connected with one's epistemological understanding and epistemic standards on what they consider strong and convincing arguments (Iordanou, 2017). Epistemological understanding follows a progression from viewing knowledge as an objective entity (Absolutists), to totally subjective (Multipists), and finally to an understanding involving coordination of the subjective and objective dimensions (Evaluativists) (Iordanou, 2017; Kuhn, 2001). Only the Evaluativist epistemological understanding provides the necessary disposition required to invest the effort to examine and evaluate alternatives, based on evidence, and reconcile diverging claims presented in multiple texts in writing.

The present study

The present research consists of an intervention study conducted in a middle school in China employing the AWM curriculum. Existing research examining the argumentative writing of Chinese students have predominantly focused on college students (e.g., Liu & Braine, 2005; Liu & Furneaux, 2015;

Wu & Rubin, 2000), particularly their argumentative writing in English as a second language (e.g., Lan et al., 2019; Pei et al., 2017). Empirical studies examining the argumentative writing of Chinese elementary or secondary school students, particularly their writing in the Chinese language, were still limited. The present study followed a quasi-experimental design with pre-assessment and post-assessment, comparing intervention and non-participating control students' individual post-assessment writing performance on the non-intervention topic of Genetically Modified Foods (GMF). As they wrote, a set of texts obtained from authentic sources and representing conflicting views on GMF were made available to each student. Following essay writing, students were asked to rank order the texts according to trustworthiness and provide written justifications. Our goal in conducting the present work was to address the following research questions:

1. Would intervention students outperform control students in integrating both position-consistent and position-inconsistent evidence from multiple texts in argumentative writing at post-assessment?
2. Were there any condition differences in students' judgment of text trustworthiness?

Method

Participants

In a quasi-experimental design, one of the school's multiple classes at the seventh grade was randomly selected to serve in the intervention condition ($n = 50$, 25 boys and 25 girls) and another to serve in the control condition ($n = 46$, 24 boys and 22 girls). The students ranged in age from 11 years 4 months to 12 years 5 months, and they all came from middle- to upper-middle-class Chinese families and spoke Chinese as their native language. The two conditions did not significantly differ in the percentage of female students and in the mean age of students. While students in the control condition participated only in the pre- and post-assessments and otherwise received their regular instruction, students in the intervention condition participated in the pre- and post-assessments, and the thrice-weekly argument curriculum (AWM) that lasted over 4 months. A total of 27 sessions, nine for each topic, were devoted to the AWM intervention.

Located in Western China, the school was a private, selective middle school, admitting roughly 30% of the students who applied. The school's curriculum primarily emphasized mastery of subject content knowledge and preparation for high-stakes, standardized tests. Classroom teaching was mostly teacher-centered and expository, with teacher-student interaction mostly following the IRE pattern (i.e., *initiation-response-evaluation*) (Mehen, 1979; Sinclair & Coulthard, 1975). Peer-to-peer dialogic argumentation, as promoted in the present AWM intervention, were largely absent from regular classroom activities at the school.

Informed consent was obtained before the intervention from all participating students and their parents, as well as from the participating teachers. Everyone involved was informed of their right to withdraw from the study at any time, but none did. They were also informed that the reporting of the study would be anonymized, and that none of the data would be released to the public. Data were collected in accordance with the standards and guidelines of the human subjects review board at Teachers College, Columbia University.

Procedure

“AWM” intervention. Three topics were addressed during the intervention (Topic 1: Should teenagers over 16 focus on their schoolwork or should they take on a part-time job? Topic 2: In order to better treat human illnesses, should animals be used to test new medical products and procedures? Topic 3: Should the sale of kidneys be legalized in China?). Each topic followed the activity sequence described below.

Session 1. Following a brief introduction to the topic, students assembled into same-side small groups of four to five based on their preferred side and each group generated reasons to support their position, recording them on index cards and sharing them with the rest of the group.

Sessions 2–6. Same-side, same-gender dyads were formed who remained together throughout these five sessions. At each session, the pair engaged in an electronic dialogue via instant-messaging software with a succession of opposing-side pairs, a different opposing-side pair in each session. While awaiting response from the opposing-side pair, dyads were asked to work on reflection sheets designed to promote reflection on the dialogues. In addition, during each session, dyads were provided on a small index card a short piece of information in the Q&A format.

Session 7. For each round of this showdown session, each side chose one member at a time to be in the “hot seat” to verbally debate a classmate from the opposing side in front of the class. Time was called after 3 min and teammates chosen by each side replaced those in the “hot seat.”

Session 8. The showdown debate was video-recorded and subsequently transcribed by the lead teacher to generate an argument map—a verbatim written record of students’ verbal exchanges in Session 7. The whole-class debrief was led by the lead teacher and students were guided through the argument map, with points awarded for effective argumentative moves and points subtracted for ineffective moves.

Session 9. The final activity for each topic cycle was students’ individual written composition of a “letter to a newspaper editor” on the topic. Students were told that the goal of their writing was to persuade readers to trust and accept their position.

Post-assessment. Once the intervention was completed, the post-assessment task was carried out in the following week. Intervention and control students were tested separately in a whole-class setting. Two consecutive class periods (80 min in total) were allocated for students to write an

argumentative essay on the following topic, “Should the Chinese government encourage or ban the production and sale of genetically modified foods?” Students were asked to first take a side on the issue and then write an argumentative essay to persuade others who disagreed with their position. The topic of GMF was not addressed during the intervention, nor was it taught in the school’s regular curriculum up to this grade level. The topic of GMF was selected because it remains a hotly debated socio-scientific controversy (Zeidler & Nichols, 2009) in the public sphere, with an abundance of contradictory and inconclusive evidence bearing on it.

An information packet with six texts was provided to each student with the following prompt, “Here are some articles related to the topic. Feel free to read them and use the information when you write. Your goal is to persuade someone who disagreed with you.” Once students completed writing, their essays were collected and students were asked to “rank order the texts from the most trustworthy to the least trustworthy and explain why you ranked the texts in that order.” The six texts remained available to students as they completed the ranking task.

The texts provided to students represented diverse, authentic source materials they might encounter when searching for information on this topic. Each text was printed on a separate sheet of paper and source information, including the name of the author or publisher, type of text, date of publication, and website (if available), was presented in the upper right corner of each text. Table 1 presents an overview of the six texts. While Texts 1, 3, and 6 supported GMF,

Table 1. Description of texts.

No.	Type of text	Publisher	Author	Position on GMF and content
1	High school textbook	People’s Education Press	Science teacher	Support: genetic engineering in agriculture
2	NGO website page	Center for Food Safety USA	Journalist	Oppose: risks of GMF for consumers
3	Government website page	Ministry of Agriculture and Rural Affairs of China	Government official	Support: safety of transgenic breeding
4	Abstract of academic journal article	<i>Critical Reviews in Foods Science and Nutrition</i>	Science researcher	Oppose: health risks of GMF
5	Social media post	Zhihu (similar to Quora)	Anonymous individual	Oppose: risks and costs of GMF
6	Corporate website page	Biocentury Transgene China	Private corporation employee	Support: advantages of saline-alkali tolerant rice

Texts 2, 4, and 5 opposed GMF. Before administering the post-assessment, the research team consulted several school staff to ensure that the texts were appropriate in terms of vocabulary and sentence structure for this cohort of students.

Coding

Coding of essays

Functional evidence-based unit. Each essay was first segmented into idea units, with a unit defined as a claim with supporting argument or evidence. An idea unit was further coded as evidence-based if it is supported by evidence. For an idea to be sufficiently supported, the selected evidence must be linked to it clearly and explicitly enough for the logical relation between the two to be revealed (Hemberger et al., 2017). If the logical relation was specified, that unit was coded as functional evidence-based unit; if not, the unit was coded as non-functional evidence-based unit.

Argumentative function. Each functional evidence-based unit was further categorized into a position-consistent or position-inconsistent unit based on its argumentative function (Shi, 2019). A position-consistent unit works in one's favor as it supports one's position or weakens an opposing position; a position-inconsistent unit works against one's favor as it weakens one's position or supports an opposing position. Also included was a super-category of evidence-based *However* unit that consists of two adjacent evidence-based units serving opposing functions and connected to one another, usually with a position-inconsistent unit followed by a position-consistent unit, indicating students' successful integration of two pieces of conflicting evidence.

Blind to condition, the first author and a Chinese-speaking colleague not involved in the present investigation coded a randomly selected 30% of the essay dataset. The two coders segmented each essay, reached an agreement of 87% and in assigning each idea unit to one of four categories (i.e., non-evidence-based, non-functional, position-consistent, and position-inconsistent), the two coders achieved an inter-rater agreement of 90.91%, Cohen's kappa = .873, $p < .0005$. All the differences in coding were resolved through discussion and the first author proceeded to code the remaining essays.

Source of evidence. We also assigned each functional evidence-based unit into one of three sources—Added, Borrowed, or Transformed—based on a modification of the coding scheme originally developed in Wiley and Voss (1996, 1999). As illustrated in Table 2, while Added evidence was drawn from students' personal knowledge, Borrowed evidence and Transformed evidence were taken from the source text(s). Borrowed evidence represented literal use of evidence in the form of copying or paraphrasing the original information, and Transformed evidence represented integrative use of information in the form of connecting it with a novel claim (i.e., integrating evidence with one's prior knowledge), or integrative use of evidence from one text or several texts that

Table 2. Coding scheme of the source of evidence.

Source of evidence	Definition
Added	Units containing novel evidence from outside the source text(s)
Borrowed	Units containing evidence taken directly or paraphrased from the source text(s)
Transformed	Units containing evidence in combination with a new claim or fact not from the source text(s) Intra-textual transformation: integrated two or more pieces of evidence from a single text that were not connected in the source text Inter-textual transformation: integrated two or more pieces of evidence from more than one source text

Source. Adapted from Wiley and Voss (1999).

were not connected in the original text(s). It is necessary to discriminate between Borrowed evidence and Transformed evidence, as the latter indicated a more advanced level of evidence use.

Two coders independently coded the same set of essays used earlier and assigned each functional evidence-based unit to one of three categories (i.e., Added, Borrowed, and Transformed), achieving an inter-rater agreement of 88.64%, Cohen's kappa = .855, $p < .0005$. Differences were resolved through discussion and the first author proceeded to code the remaining dataset.

Coding of rank justifications. Next, we coded students' written justifications of their trustworthiness ranking. To develop a coding scheme, two coders examined a portion of the dataset and segmented students' statements into idea units and independently analyzed these units openly, looking for recurrent themes. The two coders then discussed and summarized the themes into codes that were then applied to analyze more statements to check their applicability. New codes were added if new themes were identified. This iterative process of refining the coding scheme continued for several rounds until no new codes emerged.

The complete coding scheme is presented in Table 3. *Source information* referred to mention of author or publisher of the text, and the time of publication. *Content information* referred to mention of the information acquisition process, such as whether it was obtained through scientific procedures or personal speculations. The category also included other aspects of the content, including its rhetorical or logical features, as well as the text's position on GMF. *Own information* referred to mention of one's own position or background knowledge on GMF. Note that each student could mention more than one justificatory criterion. Working with 30% of the dataset, the two coders independently assigned each response to one of the ten sub-categories, achieving an inter-rater agreement of 100%. The first author proceeded to code the remaining dataset.

Table 3. Coding scheme of students' rank justifications.

Category	Sub-category	Description	Verbatim example
Source information	Author or publisher	Author or publisher of the text	A text issued by the government is more trustworthy than a social media post written by an anonymous individual.
	Date of publication	Time when the text was published	The more recent the better, because science is developing rapidly.
Content information	Information acquisition process	How information was obtained	I rated texts based on scientific procedures with higher credibility than texts based on personal speculations.
	Logic	Clarity of logic	A text is more credible if its logic is clear.
	Rhetoric	Phrasing or other rhetorical features	A text is more trustworthy if it is carefully phrased.
	Comprehensibility	Comprehensibility of information for readers	My judgment was based on the extent to which I could understand the information.
	Applicability	Applicability of information to real life context	I rated a text that is widely applicable to be more trustworthy.
	Text's position on GMF	Position of text on GMF	This text is not credible because it only talks about the advantages of GMF.
Own information	Own position on GMF	One's position on GMF	This text is more credible because it helped me deliberate on my own position.
	Background knowledge	One's prior knowledge on GMF	I used my background knowledge gained from news watching to judge the credibility of the information.

Results

Pre-assessment

To establish equivalence between the two conditions prior to the intervention, both pre-assessment of students' argumentative writing and the school's diagnostic test administered right before the start of the intervention were used. The school's diagnostic test was administered to all the seventh-

grade students during the first week of the Fall term. The goal was to provide baseline information for teachers regarding students' academic performance upon entering middle school. The essay component asked students to write a 600-word essay about a memorable personal experience. The reading comprehension component asked students to read three narratives and following each narrative, there were about five questions that assessed students' understanding and interpretation of the texts. An independent-samples t -test showed no significant condition difference in the mean score of Chinese Language Arts, $t(94) = 1.07, p = .289$, its essay component, $t(94) = -1.04, p = .303$, and its reading comprehension component, $t(94) = -.83, p = .412$.

For the pre-assessment, both intervention and control students were asked to write an argumentative essay on the following topic: Should juveniles who have committed serious crimes tried in an adult court or a juvenile court? As they write, a list of 11 short pieces of evidence in the Q&A format was distributed to each student. A negative binomial regression with condition as the predictor variable was carried out, showing no condition difference in student essays in terms of the mean number of idea units, position-consistent evidence-based units, position-inconsistent evidence-based units, and evidence-based *However* units. Detailed explanation of each of these codes is provided in the "Coding of essays" section, as provided above.

Post-assessment

Having established equivalence in performance between the two conditions at pre-assessment, we now focus on comparing their post-assessment performance to see whether there was a condition difference in their integration of textual evidence inconsistent with their own position from multiple texts and in their judgment of text trustworthiness.

Position on genetically modified foods. While 72% of intervention students and 78% of control students supported GMF, 28% of intervention students and 22% of control students opposed GMF. A chi-square test of independence indicated that there is no statistically significant association between condition and student position on GMF, $\chi^2(1, N=96) = .501, p = .479$.

Idea unit. The mean number of idea units was 9.22 ($SD = 2.01$) for the intervention condition and 8.78 ($SD = 2.59$) for the control condition, a non-significant condition difference as indicated by the Generalized Linear Model (GLM) with Poisson regression, $p = .476$. Therefore, students from the two conditions wrote essays of comparable length.

Functional evidence-based unit. The mean number of functional evidence-based units was 6.98 units ($SD = 2.00$) for the intervention condition and 5.43 units ($SD = 1.88$) for the control condition. Intervention students generated an expected 1.28 (95% CI: 1.09, 1.51) times more functional

evidence-based units than control students, a significant condition difference, Wald $\chi^2(1) = 9.12$, $p = .003$. Moreover, every intervention and control student employed functional evidence-based unit at least once.

Argumentative function. As shown in Table 4, further analyses of functional evidence-based units in terms of their argumentative function indicated that intervention and control students generated about five position-consistent units and no significant condition difference was observed. Moreover, all the students made position-consistent unit at least once. In contrast, intervention students generated slightly less than two position-inconsistent units, compared to control students who generated less than one unit. Intervention students generated an expected 2.85 (95% CI: 1.90, 4.27) times more position-inconsistent units than control students, a significant condition difference, Wald $\chi^2(1) = 25.69$, $p < .0005$. In addition, nearly 90% of intervention students made position-inconsistent unit at least once, compared to 40% of control students, a significant condition difference, $p < .0005$ (Fisher's exact test).

Moreover, intervention students generated an expected 2.12 (95% CI: 1.34, 3.36) times more evidence-based *However* units than control students, a significant condition difference, Wald $\chi^2(1) = 10.28$, $p = .001$. In addition, nearly 90% of intervention students made evidence-based *However* unit at least once, compared to 30% of control students, a significant condition difference, $p < .0005$ (Fisher's exact test).

Table 4. Mean number of different types of evidence-based claims and percentage of students employing a type at least once by condition.

Argumentative function	Source of evidence	Intervention condition ($n = 50$)		Control condition ($n = 46$)	
		M (SD)	%	M (SD)	%
Position-consistent unit	Borrowed	1.82 (0.98)	94.0	1.48 (1.49)	71.7
	Added	0.94 (0.87)	62.0	0.70 (1.01)	41.3
	Transformed	2.40 (1.41)	94.0	2.50 (1.57)	91.3
	Total	5.12 (1.55)	100.0	4.67 (1.61)	100.0
Position-inconsistent unit	Borrowed	0.82 (0.63)	70.0	0.43 (0.81)	28.3
	Added	0.10 (0.30)	10.0	0.02 (0.15)	2.2
	Transformed	0.90 (0.97)	56.0	0.22 (0.59)	15.2
	Total	1.92 (1.21)	88.0	0.67 (1.06)	37.0
Evidence-based <i>However</i> unit		1.20 (0.64)	88.0	0.57 (1.00)	30.4

Source of evidence. As shown in Table 4, further analyses of functional evidence-based units in terms of the source of evidence indicated that for position-consistent units, the two conditions showed comparable performance in making Borrowed evidence ($p = .310$), Added evidence ($p = .829$), and Transformed evidence ($p = .977$), as indicated by Fisher's Exact test. In addition, while significantly more intervention than control students made Borrowed evidence at least once ($p = .005$), a comparable proportion of students made Added evidence ($p = .065$) or Transformed evidence ($p = .707$) at least once.

For position-inconsistent units, the mean number of Added evidence was comparable across conditions. However, intervention students generated an expected 1.89 (95% CI: 1.11, 3.22) times more Borrowed evidence than control students, a significant condition difference, Wald $\chi^2(1) = 5.41$, $p = .02$. Moreover, intervention students generated an expected 4.14 (95% CI: 2.09, 8.22) times more Transformed evidence than control students, a significant condition difference, Wald $\chi^2(1) = 16.51$, $p < .0005$. In addition, less than 10% of intervention or control students made Added evidence at least once, a non-significant condition difference, $p = .206$ (Fisher's exact test). However, significantly more intervention than control students made Borrowed evidence ($p < .0005$) or Transformed evidence ($p < .0005$) at least once.

Rank order of texts by trustworthiness. Next, we examined students' rank order of the six texts and for each student, we coded the two texts they placed in the first and second ranks as high trustworthiness, the two texts in the third and fourth ranks as medium trustworthiness, and the two texts in the fifth and sixth ranks as low trustworthiness. Table 5 illustrates the percentage of students who assigned each text to these three levels of trustworthiness. For both conditions, the largest

Table 5. Percentage of students judging each text with high, medium, and low trustworthiness by condition.

Text no. and source of the text	High trustworthiness		Medium trustworthiness		Low trustworthiness	
	Intervention	Control	Intervention	Control	Intervention	Control
1. Textbook	60.0	65.2	30.0	28.3	10.0	6.5
2. NGO website	12.0	26.1	30.0	50.0	58.0*	23.9
3. Government website	68.0	78.3	24.0	15.2	8.0	6.5
4. Academic journal	40.0*	8.7	50.0	39.1	10.0*	52.2
5. Social media	8.0	13.0	6.0	19.6	86.0	67.4
6. Corporate website	12.0	8.7	60.0	47.8	28.0	43.5

Note. The asterisk notation indicates that the percentage in the intervention column significantly differs from the corresponding percentage in the control column. * $p < .008$.

percentage of students judged Text 1 and Text 3 as the most trustworthy, and the largest percentage of students judged Text 5 as the least trustworthy.

Next, we conducted a Chi-square test of independence for each text to examine whether there was a significant association between condition and judgment of text trustworthiness. To account for multiple comparisons, a Bonferroni correction was made and acceptance of statistical significance was set at $p < .008$ (six chi-square tests were carried out). Our results indicated that the two conditions showed significant difference in their judgment of Text 2, $\chi^2(2, N=96) = 11.638$, $p = .003$, as well as of Text 4, $\chi^2(2, N= 96) = 24.130$, $p < .0005$. A chi-square post-hoc test indicated that significantly more intervention than control students judged Text 2, a post from an NGO website, with low trustworthiness ($p < .001$).

Significantly more intervention than control students judged Text 4, an abstract from an academic journal article, with high trustworthiness ($p < .001$), and significantly more control than intervention students judged it with low trustworthiness ($p < .0005$). Text 4 was the only primary source in the present set of texts students received. In fact, intervention students not only trusted Text 4 to a greater extent, they also employed information from Text 4 to a greater extent in writing, as 78% of intervention students, compared to 50% of control students, made use of information from Text 4 at least once, a significant condition difference, $p = .006$ (Fisher's exact test).

Justification of trustworthiness rank. More revealing, we believe, would be students' justifications of why they ranked the texts in a certain order. For each criterion in Table 4, we calculated the percentage of students who mentioned that criterion, and the results are shown in Table 6. On average, intervention students generated 3.04 ($SD = .64$) criteria, significantly more than that of control students ($M = 2.22$, $SD = 0.99$), as indicated by an independent samples t -test, $t(94) = -4.889$, $p < .0005$.

As a next step, Fisher's exact test was carried out to examine whether there was significant condition difference in the percentage students who mentioned each criterion. To account for multiple testing, the Bonferroni correction was applied and the acceptance of statistical significance was set at $p < .005$. Our results showed that significantly more intervention than control students considered the following criteria: Information acquisition process ($p = .002$), Text's position on GMF ($p = .001$), and Own position on GMF ($p = .001$).

For the Information acquisition process, students took into account the epistemological aspect regarding how information in the text was obtained, such as "I trusted the text following scientific procedures (e.g., Text 4) and did not trust the text based on personal speculations (e.g., Text 5)." For Text's position on GMF, students gave meta-level consideration to how the author's position might influence the way in which information was presented, such as "The author supported GMF so the information might be biased." For Own position on GMF, students acknowledged at the meta-level

Table 6. Percentage of students mentioning each criterion by condition.

Justification criteria	Sub-category	Intervention	Control
Source information	Author/publisher	74.0	76.1
	Date of publication	36.0	41.3
Content information	Information acquisition process	58.0*	26.1
	Logic	18.0	8.7
	Rhetoric	16.0	23.9
	Comprehensibility	2.0	2.2
	Applicability	2.0	2.2
Own information	Text's position on GMF	42.0*	10.9
	Own position on GMF	34.0*	6.5
	Background knowledge	8.0	13.0

Note. The asterisk notation indicates that the percentage in the intervention column significantly differs from the corresponding percentage in the control column. * $p < .005$.

how their own position might influence their judgment of text trustworthiness, such as “I trusted a text more if it helped me deliberate on my own position.”

Discussion

The present work fulfilled the objective of establishing the efficacy of the AWM intervention, which focused on peer-to-peer argumentative dialogues, to enhance students' integration of conflicting information from multiple texts on a controversial issue in an individual argumentative writing task at post-assessment. Focusing on adversarial peer dialogues, the present instructional focus contributed to existing research that sought to design and implement instructional interventions to promote students' integration of multiple texts in argumentative synthesis writing (Barzilai et al., 2018). In addition, by showing that the AWM intervention supported students' deployment of more effective criteria to evaluate text trustworthiness, the present work contributed to existing studies that revealed novice students' sub-optimal performance in judging text trustworthiness (e.g., Rouet et al., 1997; Wineburg, 1991).

For our first research question, when writing from multiple texts, intervention and control students wrote essays of comparable length and both groups made about five evidence-based claims consistent with their own position. However, intervention students demonstrated enhanced abilities to make integrated use of position-inconsistent information from the source text(s), including cases in which they copied or paraphrased position-inconsistent textual information (Borrowed evidence), integrated position-inconsistent textual information with their prior knowledge or integrated multiple pieces of position-inconsistent information within a text or across texts (Transformed evidence), or integrated two pieces of conflicting textual information (evidence-based *However* units).

How did participation in the AWM curriculum supported students' enhanced abilities to attend to and make integrated use of position-inconsistent information? Our findings ruled out the interpretation that attributed intervention students' superior performance to their strengthened ability to comprehend the text(s) *per se*. A comparable percentage of intervention and control students took the pro or con position on GMF, and their performance in making use of evidence consistent with their own position was strikingly similar. In other words, once intervention and control students took a side, they were equally competent in processing and making use of textual information consistent with their preferred side.

Rather, we attributed intervention students' gains in making position-inconsistent evidence to their extended dialogic experience of exchanging claims and counterclaims with opposing-side peers (Kuhn & Halpern, 2022; Kuhn et al., 2024). In the process of arguing with opponents, participants increasingly recognized the importance of supporting claims with evidence, manifested in their increased use of meta-talk to solicit and evaluate evidence from their opponents (Kuhn et al., 2013; Shi, 2020b). As the experience of using evidence to support or weaken various claims deepened (Kuhn & Moore, 2015), recognition of the need to consider and address position-inconsistent evidence in one's contemplation of controversial issues, as well as the skills to do so, was gradually developed, consolidated, and internalized. In this sense, engagement in argumentative dialogues with an opposing side not only supported the development of strategic skills in coordinating evidence with various positions—first in interpersonal dialogues and later internalized to support individual writing (Hemberger et al., 2017)—it might also prompt the development of meta-level and epistemological awareness of the need to acknowledge and address, rather than ignore or exclude, position-inconsistent evidence.

Besides this sociocultural interpretation, an alternative, cognitive interpretation (Newell et al., 2011) was that as peer dialogues continually supported the attention to and use of position-inconsistent evidence at the strategic level, students likely began to form and apply an argument schema (Anderson & Pearson, 1984; Brewer & Nakamura, 1984; Hayes, 1996; Reznitskaya et al., 2012; Wolfe et al., 2009), defined as an abstract representation of argumentative knowledge encompassing various components of a sound argument, including counterarguments and rebuttals (Kiili et al., 2020; Wolfe et al., 2009). Future studies are thus called for to distinguish between the sociocultural and cognitive interpretations, or to show that they work in tandem in supporting intervention students' superior performance in integrating conflicting information from multiple texts.

For our second research question, analyses of students' rank order of texts according to trustworthiness and their written justifications further revealed gains on the part of intervention students. Compared to control students, intervention students trusted and used evidence from the abstract of a journal article (Text 4) to a significantly greater extent, suggesting that they valued and relied more on this primary source. This finding was particularly encouraging, as prior research consistently

reported that novices who lacked domain knowledge put too much trust in their textbooks and showed little regard for texts written by persons directly involved in the events (e.g., Britt & Aglinskias, 2002; Rouet et al., 1996; Wineburg, 1991), possibly because the years of schooling have prompted them to trust the textbook as a paramount authority (Bråten et al., 2011).

We postulate that during intervention dialogues, as students repeatedly received and practiced using Q&A evidence that contained results from original scientific research, they possibly began to recognize and endorse the value of original research in enabling them to persuade their opponents. In fact, prior analyses of peer dialogues showed an increasing trend of the use of meta-talk that sought to solicit scientific data to back up claims from their opponents (Shi, 2020b). An alternative interpretation was that given intervention students' repeated exposure to original research, they were simply primed to attend to scientific research, rather than truly recognizing its value as domain experts would do. Future studies could include follow-up interviews or engage students in think-aloud protocols (Anmarkrud et al., 2014) to further shed light on their reasoning process.

Moreover, significantly more intervention than control students explicitly acknowledged that they considered the information acquisition process when judging text trustworthiness, such as "information obtained following science procedures (Text 4) was more trustworthy than those based on personal speculations (Text 5)." These considerations, pertaining to the dimension of the *nature of knowing* in Hofer and Pintrich's (1997) individual epistemology theory, were indicative of intervention students' emerging evaluativist understanding of the need to examine the credibility of knowledge claims against the knowing process. Indeed, measures of epistemological understanding of the students participating in Study 1, as reported in an earlier work (Shi, 2020b), showed intervention students' greater progression toward multiplist or evaluativist thinking, in comparison to control students who were mostly thinking in absolutist terms.

In addition, significantly more intervention than control students were metacognitively aware of the sidedness of a text or of their own position on GMF when judging text trustworthiness, possibly indicating that intervention students developed some nascent awareness that an author's position might influence how information is presented and a reader's position might influence how information is interpreted. Again, these gains could be partially attributed to the extended dialogic experience of intervention students; as they argued with opposing-side peers, conflicting positions were continually made explicit and dealt with, highlighting to students the necessity to take into consideration diverging positions on a controversial issue when processing conflicting information sources.

To take a step further, we would like to invoke Richter and Maier's (2017) two-step model of processing conflicting information in multiple documents, which specified that when encountering conflicting information, readers engage in epistemic monitoring and they tend to concentrate their cognitive resources on information consistent with prior beliefs. The elaborative processing of belief-inconsistent information would occur only when readers are motivated and cognitively

capable, as the latter was more resource demanding and under the strategic control of the reader. Applying this two-step model to our results, we propose that it is possible that a greater proportion of intervention students, being more cognizant of a conflicting position on a controversial issue, might be better supported in self-regulating their intentional engagement in the more strategic, resource-intensive elaboration of belief-inconsistent information, leading to their enhanced performance in integrating position-inconsistent evidence in argumentative writing. These findings were also consistent with Kuhn's (2019, 2022) theoretical framework on the development of argument skills, which proposes strong metacognitive competence and appropriate epistemological standards and intellectual values and dispositions as necessary conditions for the development of argument skills, particularly skills related to attending to and coordinating own and other's perspectives (Iordanou & Kuhn, 2020; Kuhn & Modrek, 2021; Kuhn & Udell, 2007).

How well students deliberate controversial socio-scientific issues, such as genetically modified foods as examined in the present study, may have far reaching implications for individual life and for society at large (Shi et al., 2021, 2024a, 2024b). In an era of information explosion enabled by continuous technological breakthroughs, it has become more imperative than ever to instill in the young citizenry the skills and dispositions to choose wisely which sources to trust and the ability to reconcile discrepant accounts in one's mental representation of and writing on an open-ended issue. Following a sociocultural approach, the present research opened promising perspectives from an instructional standpoint, that is, peer-to-peer dialogues focusing on adversarial argumentation could facilitate students' integration of conflicting information from multiple texts in a post-dialogue argumentative writing task, as well as facilitating students' enhanced judgment and differentiation of reliable information sources, from less reliable ones.

Contributorship

Yuchen Shi performed data collection and analysis and wrote the manuscript. Kalypso Iordanou contributed to designing the study, writing the Introduction section, interpreting the results, and revising the manuscript.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


Ethical statement

The work reported has not been previously published, and is not being (in present or revised form) considered for publication in other venues. Data were collected in accordance with the standards and guidelines of the human subjects review board at Teachers College, Columbia University.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Yuchen Shi  <https://orcid.org/0000-0002-1350-0843>

References

- Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. In P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 255–291). Longman, Inc.
- Anmarkrud, Ø, Bråten, I., & Strømsø, H. I. (2014). Multiple-documents literacy: Strategic processing, source awareness, and argumentation when reading multiple conflicting documents. *Learning and Individual Differences, 30*, 64–76. <https://doi.org/10.1016/j.lindif.2013.01.007>
- 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. <https://doi.org/10.1007/s11409-016-9165-7>
- Barzilai, S., Zohar, A. R., & Mor-Hagani, S. (2018). Promoting integration of multiple texts: A review of instructional approaches and practices. *Educational Psychology Review, 30*(3), 973–999. <https://doi.org/10.1007/s10648-018-9436-8>
- Billig, M. (1987). *Arguing and thinking: A rhetorical approach to social psychology*. Cambridge University Press.
- Braasch, J. L. G., & Kessler, E. D. (2021). Working toward a theoretical model for source comprehension in everyday discourse. *Discourse Processes, 58*(5-6), 449–467. <https://doi.org/10.1080/0163853X.2021.1905393>
- Bråten, I., & Braasch, J. L. G. (2018). The role of conflict in multiple source use. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 184–201). Routledge/Taylor & Francis Group.
- Bråten, I., & Strømsø, H. I. (2009). Effects of task instruction and personal epistemology on the understanding of multiple texts about climate change. *Discourse Processes, 47*(1), 1–31. <https://doi.org/10.1080/01638530902959646>
- Bråten, I., Strømsø, H. I., & Britt, M. A. (2009). Trust matters: Examining the role of source evaluation in students' construction of meaning within and across multiple texts. *Reading Research Quarterly, 44*(1), 6–28. <https://doi.org/10.1598/RRQ.44.1.1>
- Bråten, I., Strømsø, H. I., & Salmerón, L. (2011). Trust and mistrust when students read multiple information sources about climate change. *Learning and Instruction, 21*(2), 180–192. <https://doi.org/10.1016/j.learninstruc.2010.02.002>
- Brewer, W. F., & Nakamura, G. V. (1984). The nature and functions of schemas. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 1, pp. 118–160). Lawrence Erlbaum Associates Publishers.
- Britt, M. A., & Aglinskias, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction, 20*(4), 485–522. https://doi.org/10.1207/S1532690XCI2004_2

- Cazden, C. (2001). *Classroom discourse: The language of teaching and learning*. Heinemann.
- Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, 63(1), 1–49. <https://doi.org/10.3102/00346543063001001>
- Crowell, A., & Kuhn, D. (2014). Developing dialogic argumentation skills: A 3-year intervention study. *Journal of Cognition and Development*, 15(2), 363–381. <https://doi.org/10.1080/15248372.2012.725187>
- De La Paz, S., & 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. <https://doi.org/10.1016/j.cedpsych.2010.03.001>
- Du, H., & List, A. (2021). Evidence use in argument writing based on multiple texts. *Reading Research Quarterly*, 56(4), 715–735. <https://doi.org/10.1002/rrq.366>
- Gergen, K. (2015). *An invitation to social construction* (3rd ed.). Sage.
- Gil, L., Bråten, I., Vidal-Abarca, E., & Strømsø, H. I. (2010a). Summary versus argument tasks when working with multiple documents: Which is better for whom? *Contemporary Educational Psychology*, 35(3), 157–173. <https://doi.org/10.1016/j.cedpsych.2009.11.002>
- Gil, L., Bråten, I., Vidal-Abarca, E., & Strømsø, H. I. (2010b). Understanding and integrating multiple science texts: Summary tasks are sometimes better than argument tasks. *Reading Psychology*, 31(1), 30–68. <https://doi.org/10.1080/02702710902733600>
- Goldman, S. R., & Scardamalia, M. (2013). Managing, understanding, applying, and creating knowledge in the information age: Next-generation challenges and opportunities. *Cognition and Instruction*, 31(2), 255–269. <https://doi.org/10.1080/10824669.2013.773217>
- 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. <https://doi.org/10.1037/0022-0663.99.3.445>
- 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. <https://doi.org/10.1007/s11145-019-09939-6>
- Hagen, ÅM, Braasch, J. L. G., & Bråten, I. (2014). Relationships between spontaneous note-taking, self-reported strategies and comprehension when reading multiple texts in different task conditions. *Journal of Research in Reading*, 37(S1), S141–S157. <https://doi.org/10.1111/j.1467-9817.2012.01536.x>
- Hämäläinen, E. K., Kiili, C., Marttunen, M., Räikkönen, E., González-Ibáñez, R., & Leppänen, P. H. T. (2020). Promoting sixth graders' credibility evaluation of web pages: An intervention study. *Computers in Human Behavior*, 110, 106372. <https://doi.org/10.1016/j.chb.2020.106372>
- Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1–27). Lawrence Erlbaum Associates, Inc.
- Hemberger, L., Kuhn, D., Matos, F., & Shi, Y. (2017). A dialogic path to evidence-based argumentative writing. *Journal of the Learning Sciences*, 26(4), 575–607. <https://doi.org/10.1080/10508406.2017.1336714>
- Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, 67(1), 88–140. <https://doi.org/10.3102/00346543067001088>

- Iordanou, K. (2017). From theory of mind to epistemic cognition. A lifespan perspective. *Frontline Learning Research, 4*(5), 106–119. <https://doi.org/10.14786/flr.v4i5.252>
- Iordanou, K., & Kuhn, D. (2020). Contemplating the opposition: Does a personal touch matter? *Discourse Processes, 57*(4), 343–359. <https://doi.org/10.1080/0163853X.2019.1701918>
- Iordanou, K., Kuhn, D., Matos, F., Shi, Y., & Hemberger, L. (2019). Learning by arguing. *Learning and Instruction, 63*, 101207. <https://doi.org/10.1016/j.learninstruc.2019.05.004>
- Iordanou, K., & Rapanta, C. (2021). “Argue with me”: A method for developing argument skills. *Frontiers in Psychology, 12*, 631203. <https://doi.org/10.3389/fpsyg.2021.631203>
- Kienhues, D., Stadler, M., & Bromme, R. (2011). Dealing with conflicting or consistent medical information on the web: When expert information breeds laypersons’ doubts about experts. *Learning and Instruction, 21*(2), 193–204. <https://doi.org/10.1016/j.learninstruc.2010.02.004>
- Kiili, C., Bråten, I., Kullberg, N., & Leppänen, P. H. T. (2020). Investigating elementary school students’ text-based argumentation with multiple online information resources. *Computers & Education, 147*, 103785. <https://doi.org/10.1016/j.compedu.2019.103785>
- Kiili, C., Bråten, I., Strømsø, H. I., Hagerman, M. S., Rääkkönen, E., & Jyrkiäinen, A. (2022). Adolescents’ credibility justifications when evaluating online texts. *Education and Information Technologies, 27*(6), 7421–7450. <https://doi.org/10.1007/s10639-022-10907-x>
- Klaczynski, P. A., & Gordon, D. H. (1996). Self-serving influences on adolescents’ evaluations of belief-relevant evidence. *Journal of Experimental Child Psychology, 62*(3), 317–339. <https://doi.org/10.1006/jecp.1996.0033>
- Kuhn, D. (2000). Metacognitive development. *Current Directions in Psychological Science, 9*(5), 178–181. <https://doi.org/10.1111/1467-8721.00088>
- Kuhn, D. (2001). How do people know? *Psychological Science, 12*(1), 1–8. <https://doi.org/10.1111/1467-9280.00302>
- Kuhn, D. (2019). Critical thinking as discourse. *Human Development, 62*(3), 146–164. <https://doi.org/10.1159/000500171>
- Kuhn, D. (2022). Metacognition matters in many ways. *Educational Psychologist, 57*(2), 73–86. <https://doi.org/10.1080/00461520.2021.1988603>
- Kuhn, D., Bruun, S., & Geithner, C. (2024). Enriching thinking through discourse. *Cognitive Science, 48*(3), e13420. <https://doi.org/10.1111/cogs.13420>
- Kuhn, D., & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents’ thinking. *Psychological Science, 22*(4), 545–552. <https://doi.org/10.1177/0956797611402512>
- Kuhn, D., Goh, W., Iordanou, K., & Shaenfield, D. (2008). Arguing on the computer: A microgenetic study of developing argument skills in a computer-supported environment. *Child Development, 79*(5), 1310–1328. <https://doi.org/10.1111/j.1467-8624.2008.01190.x>
- Kuhn, D., & Halpern, M. (2022). How might argumentation research inform discourse-based social studies education? *The Social Studies, 113*(5), 264–270. <https://doi.org/10.1080/00377996.2022.2053832>
- Kuhn, D., Hemberger, L., & Khait, V. (2016). *Argue with me: Argument as a path to developing students’ thinking and writing* (2nd ed.). Routledge.
- Kuhn, D., & Modrek, A. (2021). Mere exposure to dialogic framing enriches argumentative thinking. *Applied Cognitive Psychology, 35*(5), 1349–1355. <https://doi.org/10.1002/acp.3862>

- Kuhn, D., & Moore, W. (2015). Argumentation as core curriculum. *Learning: Research and Practice, 1*(1), 66–78. <https://doi.org/10.1080/23735082.2015.994254>
- Kuhn, D., & Udell, W. (2007). Coordinating own and other perspectives in argument. *Thinking and Reasoning, 13*(2), 90–104. <https://doi.org/10.1080/13546780600625447>
- Kuhn, D., Zillmer, N., Crowell, A., & Zavala, J. (2013). Developing norms of argumentation: Metacognitive, epistemological, and social dimensions of developing argumentative competence. *Cognition and Instruction, 31*(4), 456–496. <https://doi.org/10.1080/07370008.2013.830618>
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480–498. <https://doi.org/10.1037/0033-2909.108.3.480>
- Lan, G., Lucas, K., & Sun, Y. (2019). Does L2 writing proficiency influence noun phrase complexity? A case analysis of argumentative essays written by Chinese students in a first-year composition course. *System, 85*, 102116. <https://doi.org/10.1016/j.system.2019.102116>
- Le Bigot, L., & Rouet, J.-F. (2007). The impact of presentation format, task assignment, and prior knowledge on students' comprehension of multiple online documents. *Journal of Literacy Research, 39*(4), 445–470. <https://doi.org/10.1080/10862960701675317>
- List, A., Alexander, P. A., & Stephens, L. A. (2017). Trust but verify: Examining the association between students' sourcing behaviors and ratings of text trustworthiness. *Discourse Processes, 54*(2), 83–104. <https://doi.org/10.1080/0163853X.2016.1174654>
- Litman, C., Marple, S., Greenleaf, C., Charney-Sirott, I., Bolz, M. J., Richardson, L. K., Hall, A. H., George, M., & Goldman, S. R. (2017). Text-based argumentation with multiple sources: A descriptive study of opportunity to learn in secondary English language arts, history, and science. *Journal of the Learning Sciences, 26*(1), 79–130. <https://doi.org/10.1080/10508406.2016.1256809>
- Liu, M., & Braine, G. (2005). Cohesive features in argumentative writing produced by Chinese undergraduates. *System, 33*(4), 623–636. <https://doi.org/10.1016/j.system.2005.02.002>
- Liu, X., & Furneaux, C. (2015). Argument structures in Chinese university students' argumentative writing: A contrastive study. *English Text Construction, 8*(1), 65–87. <https://doi.org/10.1075/etc.8.1.03liu>
- 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. <https://doi.org/10.1007/s11409-013-9111-x>
- 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. <https://doi.org/10.1080/07370008.2018.1425300>
- Matos, F. (2021). Collaborative writing as a bridge from peer discourse to individual argumentative writing. *Reading and Writing: An Interdisciplinary Journal, 34*(5), 1321–1342. <https://doi.org/10.1007/s11145-020-10117-2>
- Mead, G. H. (1934). *Mind, self, and society*. University of Chicago Press.
- Mehen, 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*(2), 57–74. <https://doi.org/10.1017/S0140525X10000968>
- 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. <https://doi.org/10.1598/RRQ.46.3.4>

- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175–220. <https://doi.org/10.1037/1089-2680.2.2.175>
- Nussbaum, E. M. (2008). Using argumentation vee diagrams (AVDs) for promoting argument-counterargument integration in reflective writing. *Journal of Educational Psychology, 100*(3), 549–565. <https://doi.org/10.1037/0022-0663.100.3.549>
- Nussbaum, E. M., & Schraw, G. (2007). Promoting argument-counterargument integration in students' writing. *Journal of Experimental Education, 76*(1), 59–92. <https://doi.org/10.3200/JEXE.76.1.59-92>
- O'Keefe, D. J. (1999). How to handle opposing arguments in persuasive messages: A meta-analytic review of the effects of one-sided and two-sided messages. *Annals of the International Communication Association, 22*(1), 209–249. <https://doi.org/10.1080/23808985.1999.11678963>
- Papathomas, L., & Kuhn, D. (2017). Learning to argue via apprenticeship. *Journal of Experimental Child Psychology, 159*, 129–139. <https://doi.org/10.1016/j.jecp.2017.01.013>
- Pei, Z., Zheng, C., Zhang, M., & Liu, F. (2017). Critical thinking and argumentative writing: Inspecting the association among EFL learners in China. *English Language Teaching, 10*(10), 31–42. <https://doi.org/10.5539/elt.v10n10p31>
- Rapanta, C. (2021). Can teachers implement a student-centered dialogical argumentation method across the curriculum? *Teaching and Teacher Education, 105*, 103404. <https://doi.org/10.1016/j.tate.2021.103404>
- Rapanta, C., & Felton, M. K. (2022). Learning to argue through dialogue: A review of instructional approaches. *Educational Psychology Review, 34*(2), 477–509. <https://doi.org/10.1007/s10648-021-09637-2>
- Resnick, L., Asterhan, C., & Clarke, S. (2015). *Socializing intelligence through academic talk and dialogue*. American Educational Research Association.
- Reznitskaya, A., Anderson, R. C., McNurlen, B., Nguyen-Jahiel, K., Archodidou, A., & Kim, S.-Y. (2001). Influence of oral discussion on written argument. *Discourse Processes, 32*(2-3), 155–175. <https://doi.org/10.1080/0163853X.2001.9651596>
- Reznitskaya, A., Glina, M., Carolan, B., Michaud, O., Rogers, J., & Sequeira, L. (2012). Examining transfer effects from dialogic discussions to new tasks and contexts. *Contemporary Educational Psychology, 37*(4), 288–306. <https://doi.org/10.1016/j.cedpsych.2012.02.003>
- Reznitskaya, A., Kuo, L. J., Clark, A. M., Miller, B., Jadallah, M., Anderson, R. C., & Nguyen-Jahiel, K. (2009). Collaborative reasoning: A dialogic approach to group discussions. *Cambridge Journal of Education, 39*(1), 29–48. <https://doi.org/10.1080/03057640802701952>
- Richter, T., & Maier, J. (2017). Comprehension of multiple documents with conflicting information: A two-step model of validation. *Educational Psychologist, 52*(3), 148–166. <https://doi.org/10.1080/00461520.2017.1322968>
- Rouet, J.-F., Britt, M. A., Mason, R. A., & Perfetti, C. A. (1996). Using multiple sources of evidence to reason about history. *Journal of Educational Psychology, 88*(3), 478–493. <https://doi.org/10.1037/0022-0663.88.3.478>
- Rouet, J.-F., Favart, M., Britt, M. A., & Perfetti, C. A. (1997). Studying and using multiple documents in history: Effects of discipline expertise. *Cognition and Instruction, 15*(1), 85–106. https://doi.org/10.1207/s1532690xci1501_3
- Rouet, J.-F., & Potocki, A. (2018). From reading comprehension to document literacy: Learning to search for, evaluate and integrate information across texts/De la lectura a la alfabetización documental: Aprender a

- buscar, evaluar e integrar información de diversos textos. *Journal for the Study of Education and Development*, 41(3), 415–446. <https://doi.org/10.1080/02103702.2018.1480313>
- Shi, Y. (2019). Enhancing evidence-based argumentation in a Mainland China middle school. *Contemporary Educational Psychology*, 59, 101809. <https://doi.org/10.1016/j.cedpsych.2019.101809>
- Shi, Y. (2020a). Constructed dialogs reveal skill development in argumentative writing. *Reading and Writing*, 33(9), 2311–2335. <https://doi.org/10.1007/s11145-020-10045-1>
- Shi, Y. (2020b). Talk about evidence during argumentation. *Discourse Processes*, 57(9), 770–792. <https://doi.org/10.1080/0163853X.2020.1777498>
- Shi, Y. (2024). The effects of discourse goals on written arguments in elementary school students. *International Journal of Educational Research*, 126, 102375. <https://doi.org/10.1016/j.ijer.2024.102375>
- 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, 100533. <https://doi.org/10.1016/j.lcsi.2021.100533>
- Shi, Y., Zhang, Z., & Cao, S. (2024a). A case study of emergent leadership in whole-class discussion of controversial issues. *Language and Education*. <https://doi.org/10.1080/09500782.2024.2391296>
- Shi, Y., Zhang, Z., Cao, S., & Liu, Q. (2024b). Dialogic teaching of controversial issues: Discursive moves to enact two-sided discussions. *Language and Education*, 38(2), 303–319. <https://doi.org/10.1080/09500782.2023.2240292>
- Sinatra, G. M., & Lombardi, D. (2020). Evaluating sources of scientific evidence and claims in the post-truth era may require reappraising plausibility judgments. *Educational Psychologist*, 55(3), 120–131. <https://doi.org/10.1080/00461520.2020.1730181>
- Sinclair, J. M. H., & Coulthard, M. (1975). *Towards an analysis of discourse: The English used by pupils and teachers*. Oxford University Press.
- Stahl, S. A., Hynd, C. R., Britton, B. K., McNish, M. M., & Bosquet, D. (1996). What happens when students read multiple source documents in history? *Reading Research Quarterly*, 31(4), 430–456. <https://doi.org/10.1598/RRQ.31.4.5>
- Strømso, H. I., Bråten, I., & Britt, M. A. (2011). Do students' beliefs about knowledge and knowing predict their judgement of texts' trustworthiness? *Educational Psychology*, 31(2), 177–206. <https://doi.org/10.1080/01443410.2010.538039>
- 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. <https://doi.org/10.1007/BF00128147>
- Wertsch, J. V. (1979). From social interaction to higher psychological processes: A clarification and application of Vygotsky's theory. *Human Development*, 22(1), 1–22. <https://doi.org/10.1159/000272425>
- Wiley, J., & Voss, J. F. (1996). The effects of 'playing historian' on learning in history. *Applied Cognitive Psychology*, 10(7), 63–72. [https://doi.org/10.1002/\(SICI\)1099-0720\(199611\)10:7<63::AID-ACP438>3.0.CO;2-5](https://doi.org/10.1002/(SICI)1099-0720(199611)10:7<63::AID-ACP438>3.0.CO;2-5)
- Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology*, 91(2), 301–311. <https://doi.org/10.1037/0022-0663.91.2.301>

- Wineburg, S. S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology, 83*(1), 73–87. <https://doi.org/10.1037/0022-0663.83.1.73>
- 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. <https://doi.org/10.1037/edu0000043>
- Wolfe, C. R. (2011). Argumentation across the curriculum. *Written Communication, 28*(2), 193–219. <https://doi.org/10.1177/0741088311399236>
- Wolfe, C. R., & Britt, M. A. (2008). The locus of the myside bias in written argumentation. *Thinking & Reasoning, 14*(1), 1–27. <https://doi.org/10.1080/13546780701527674>
- 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. <https://doi.org/10.1177/0741088309333019>
- Wu, S.-Y., & Rubin, D. L. (2000). Evaluating the impact of collectivism and individualism on argumentative writing by Chinese and North American college students. *Research in the Teaching of English, 35*(2), 148–178. <https://doi.org/10.58680/rte20001715>
- Zeidler, D. L., & Nichols, B. H. (2009). Socioscientific issues: Theory and practice. *Journal of Elementary Science Education, 21*(2), 49. <https://doi.org/10.1007/BF03173684>