



# Supporting strategic and meta-strategic development of argument skill: the role of reflection

Kalypso Iordanou<sup>1</sup>

Received: 8 June 2020 / Accepted: 19 December 2021  
© The Author(s) 2022

## Abstract

The present work examines the role of reflection in supporting the development of argument skill. Participants who engaged in argumentation practice with additional reflective activities outperformed a control group who only engaged in the argumentation practice. The experimental group showed greater gains in developing argument skill – particularly in employing evidence to weaken an opposing position. Experimental condition participants also exhibited greater gains in content knowledge on the intervention topic compared to the control condition participants. Microgenetic analysis of dialogs during the interventions revealed a different pattern of progress across the two conditions. Experimental condition participants exhibited gradual and overall greater improvements at both strategic and meta levels – meta-strategic and epistemic – compared to control condition participants. Findings support the hypothesis that engagement in reflection during dialogic argumentation supports the development of metacognition, both as competence and disposition, which in turn supports performance at the strategic level. Educational implications are discussed, stressing the beneficial role of engagement in reflective activities for promoting argument skill, above and beyond practice.

**Keywords** Argumentation · Metacognition · Epistemic · Development · Learning · Argument skill

The idea that argument skill is vital to critical thinking for human beings is not new; it goes back to Plato and Socrates. Argument skill is not only fundamental for social interactions, it is integrated into one's own thinking processes; "Thought and speech are the same; only the former, which is a silent inner conversation of the soul with itself, has been given the special name of thought" (Plato's dialogue *Sophist*, cited by Billig, 1987, p.111). The need to develop argument skill remains highly topical. Polarization of attitudes and beliefs is increasing, in an era when access to diverse sources and viewpoints has become easier than ever before. Individuals become echo chambers of their own views, without questioning the foundations of their views or considering alternatives (Kuhn & Iordanou, [in press](#)). There is clearly a need to develop individuals' argument skill, that involves the ability to

---

✉ Kalypso Iordanou  
KIordanou@uclan.ac.uk

<sup>1</sup> University of Central Lancashire, 12 – 14 University Avenue Pyla, 7080 Larnaka, Cyprus

critically evaluate a claim and the reasons given in its support (Macagno et al., 2018), and construct counterarguments grounded on evidence (Toulmin, 1958). The development of argument skill is a fundamental objective of educational curricula (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010; NGSS Lead States, 2013). While there is agreement among psychologists, philosophers and policymakers regarding the need to support the development of individuals' argument skill, we lack comprehensive understanding of how we can achieve this objective. Even less clear is our understanding of the mechanism that supports the development of argument skill. How does argument skill develop, and how can we support individuals' development of this skill?

The focus of recent research aiming to support argument skill is to utilize scaffolds such as visual representations (Dwyer et al., 2012; Nussbaum et al., 2019), or to develop computer software (Clark et al., 2007). Other researchers focus on prompting reflection using teachers' questions as a tool (Reznitskaya et al., 2012), or role-playing, by asking students to take on the role of an evaluator of others' dialogs (Felton, 2004; Lin et al., 2012). The implementation of some intervention programs provides some evidence of success in improving individuals' argument skill, yet those studies do not fully illuminate our understanding of how development occurred, and more specifically what triggered it. The educational programs applied in classrooms are usually complex and involve many elements, therefore the contribution of the individual elements of a curriculum in supporting the development of argument skill is not clear. For example, there is substantial empirical evidence that the "Argue with me" dialog-based pedagogical method (Kuhn et al., 2016) supports the ability to construct evidence-based arguments to address opposing positions (two-sided arguments) (see Iordanou & Rapanta, 2021, for a review). It involves different activities, such as practice in dialogic argumentation – through engagement in a series of dyadic interactions with multiple partners –, collaborative work in pairs and small groups, and engagement in reflective activities. The present work aims to provide insight into the mechanism question by examining the role of *reflection* in supporting argument skill development. Although the significance of reflection on one's own thought, where one treats one's own thought as itself an object of examination, has been noted by several philosophers, educators, and psychologists, we have yet to appreciate fully what that significance is (Kuhn, 1991).

Empirical studies show that engagement in argumentative practice and reflection supports strategic development, such as construction of counterarguments, although the question of how this happens remains an open one. Felton (2004) compared adolescents who engaged in verbal discussion, and then in reflection with a peer who did not engage in the discussion, with adolescents who engaged only in discussion, found that the former group showed greater gains in the use of strategies associated with advances in argument skill, particularly in a transfer topic. Iordanou and Constantinou (2015) introduced reflective activities focusing specifically on the use of evidence in argumentation. In that study students who engaged in argumentation and evidence-focused reflective activities, in the context of a learning environment, increased the use of evidence in their dialogs, used more evidence that functioned to weaken opponents' claims and used more accurate evidence, compared to a group of students who used the same learning environment but did not engage in an argumentative discourse activity. Shi (2019) found that adding evidence-focused reflective activities as an additional component in the "Argue with me" curriculum benefited the construction of evidence-based arguments, compared to engagement in the same curriculum without this extra component, or in the regular school curriculum. Previous studies have not, however, examined *whether* and, most importantly, *how* reflection supports use

of evidence in arguments above and beyond dialogic practice, which is the focus of the present study.

The hypothesis examined here is that engagement in reflection while individuals are engaged in dialogic argumentation will support the development of meta-level understanding of the goals of skilled argumentation, namely understanding of counterargument and use of evidence as objectives of skilled argumentation (Walton, 1989). This hypothesis is examined using a microgenetic method. Previous work involving the “Argue with me” curriculum and employing the microgenetic method showed that meta level understanding was developing (Kuhn et al., 2013, 2008) along with strategic development, that is, construction of counterarguments and rebuttals. Yet, the question of how this development happened, in other words, which element or elements of the curriculum might have triggered it, remains an open one. The present study addresses this question by putting reflection under rigorous experimental investigation. We extend previous studies which either did not employ a control group (Kuhn et al., 2008), or employed control groups that involved either engagement in a non-argumentative activity (Iordanou & Constantinou, 2015) or a combination of argumentative and reflective activities (Shi, 2019), by employing a control group that engaged only in argumentative activities.

To examine the role of engagement in reflective activities in supporting the development of argument skill, we employed a control group where the only difference between the experimental and control group was the engagement in additional reflective activities by experimental group participants. Support for this design comes from the study by Felton (2004), which found that a condition in which adolescents engaged in reflective activities with peer advisors who observed and then discussed possible improvements with them, was more effective in supporting argument skill relative to a condition in which participants argued without peer advisors. In the present study, participants engaged in reflection on their own dialog, using a reflection sheet and a transcript of their dialog as the basis for this activity. This is different from previous studies where reflection was supported by an “external” individual (Felton, 2004; Lin et al., 2012; Reznitskaya et al., 2012). The reflection sheet prompted participants to reflect on their response to others’ arguments and counterarguments, as the advisors did in Felton’s (2004) study. In addition, the reflection sheet asked participants to also reflect on the use of evidence in their response. A particular focus of the present work is using the microgenetic method to examine the process of change of both the experimental and control condition participants while they were engaged in dialogic activities in the context of an intervention. This is the first time, to the best of my knowledge, where the process of change – at both strategic and meta-strategic level – is examined and contrasted between a condition which engaged in argumentative and reflective activities and a condition which engaged only in argumentative activities. The present study has the potential to provide insights to open questions in the literature regarding strategic and meta-strategic development and the conditions under which it takes place (Veenman et al., 2006), examining in particular the role of reflection in developing meta-level understanding of argumentation, as well as developing strategic skill to execute it.

## Theoretical background

Metacognition is defined as the knowledge about, monitoring and control of one’s own cognitive processes (Flavell, 1979). Since Flavell introduced the construct, four decades ago, metacognition has received prominent attention in psychology. Developmental studies

showed that metacognition develops over the first decades of life (Roebbers, 2017; Schneider & Löffler, 2016), but our understanding of the trajectories of this development is incomplete, with even adults reported as failing to have good judgments of their own knowing (Finn & Metcalfe, 2014). The question of how to support the development of metacognition is a pressing one, yet our understanding of how to address this is still limited. Metacognition can be conceptualized both as a competence and as a disposition (Kuhn, 2001). Metacognition as a competence refers to one's knowledge of strategies and the ability to control their effective application in the context of a particular task in order to achieve specific goals (Barzilai, & Ka'adan, 2017; Flavell, 1979; Metcalfe, 2009). Metacognition as a disposition refers to epistemic understanding of the point of an intellectual task, that is, valuing an intellectual task as a practice on which it is worthy to expend cognitive effort to engage in, and is closely connected with one's motivation to engage in an intellectual task (Chinn et al., 2020; Efklides, 2011; Iordanou et al., 2019b; Kuhn, 2020; Mason & Scirica, 2006).

### Metacognition: meta-strategic knowledge

Metacognition as a competence involves meta-strategic knowledge, that is, awareness and understanding of the strategies available in one's repertoire that are potentially applicable to a task, as well as of their power in relation to the task objectives (Zohar & Ben David, 2009) and control of the application of strategies (Iordanou, 2016a). It has been proposed that meta-strategic knowledge is fundamental for strategic performance (Barzilai & Zohar, 2012; Kuhn, 2000). For example, a relation has been reported between meta-strategic knowledge about memory strategies and the memory strategies employed (Schneider & Pressley, 1997).

The question of how meta-strategic knowledge can be supported is still an open one. Some researchers have proposed that explicit knowledge about cognitive processes, what Flavell and Wellman (1975) called declarative knowledge about strategies, relates to procedural metacognition and task performance (Nelson & Narens, 1990). Supporting this position are empirical findings showing that explicit teaching of meta-strategic knowledge has a positive effect on participants' meta-strategic knowledge and the application of strategies, such as the control of variables strategy (Zohar & Ben David, 2009). Yet other researchers, who did not find a relation between the declarative and the procedural components of metacognition, have proposed that metacognitive abilities are directed by *implicit* meta-level knowledge, which is informed by one's experience with engagement in cognitive processes, rather than by the explicit declarative knowledge that can be verbalized – the theory of dual system of metacognitive abilities (Tsalas et al., 2017). Engagement in tasks that involve targeted strategies is therefore pivotal for the development of meta-level knowledge of those strategies and the effective usage of strategies at the cognitive level (Schneider & Pressley, 1997; Tsalas et al., 2017). The application of strategies at the cognitive level informs meta-strategic knowing at the meta-level, which at the same time directs the application of strategies in a continuous cycle in which the meta-level both influences and is influenced by the cognitive level (Kuhn, 2000).

In the present work, the role of reflection on the argumentation process is examined as a means to promote meta-strategic knowledge of argument skills. It is hypothesized that reflecting on one's strategies with the aim to judge their effectiveness in relation to the task's goal, that in the context of argumentation is to persuade others, enhances one's meta-strategic knowledge. We further hypothesize that the dialogic argumentative context

constitutes a facilitative condition for reflection, compared with solitary conditions where one needs to imagine how a strategy might be perceived by another person. In contrast, in argumentation, others' reactions to a particular argument we have presented are readily available in our memory, in the case of verbal communication, or in a transcript, in the case of remote communication. Besides convenience, the dialogic argumentative context provides more accurate information regarding the persuasive power of a strategy, that is fundamental for informing meta-strategic knowing, by receiving feedback from a real person in the context of an authentic discussion, compared to imagining how it might be received in the context of a hypothetical discussion.

### **Metacognition: epistemic understanding**

Another form of metacognition that directs performance at the cognitive level is epistemic understanding (Bråten et al., 2011; Chinn et al., 2020; Kuhn, 2020; Kuhn et al., 2013; Wiley et al., 2020). Whether an individual will pay attention to others' positions and engage in critical evaluation of evidence that supports positions, depends not only on whether individuals have the knowledge and skills to do so – including both strategic and meta-strategic knowledge – but also on whether individuals have the disposition to engage in this task. The latter depends on one's epistemic understanding. Epistemic understanding undergo development throughout the lifespan (King & Kitchener, 1994; Perry, 1970). The roots of epistemic understanding can be traced in the early conception of evidence as an objective reality, through achievements in what are known as Theory of Mind tasks (Iordanou, 2016a). The first developmental level of epistemic understanding is Absolutism. Absolutists believe that truth is objective, and reality speaks for itself. Individuals are either right or wrong. In the case of disagreement, an alternative claim will be dismissed as a false belief requiring correction, given an individual's belief in a single objective reality.

Multiplism is the second developmental level of epistemic understanding. The understanding in early adolescence that evidence is amenable to different interpretations which give rise to different beliefs and theories, signals the precursor of the acknowledgment of the interpretive nature of the human mind and the subjective nature of knowledge (Lalonde & Chandler, 2002). Multiplists acknowledge that knowledge is subjective and that multiple interpretations or opinions may exist on a particular issue. At this stage of development, in the case of disagreement with another person on a particular matter, individuals will agree that they disagree and they don't see a point in engaging in a discussion. The final level of development is the Evaluativist level, where the subjective and objective dimensions of knowing are coordinated. Knowledge is viewed as judgment, which is formed after careful evaluation of alternative positions, aiming to find the best explanatory framework which is supported by the data available at the time that the judgment is made. Evaluativists view beliefs as judgments that are amenable to change in the light of new evidence, or of an alternative theory with a greater explanatory power. Therefore, in contrast to absolutists and multiplists, evaluativists view argumentation as a valuable activity to engage in to help them form a judgment on an issue.

Epistemic understanding is not only connected with one's attitude towards argumentation, influencing whether one will engage in it or not, but also with one's performance during argumentation; that is, how one argues. Epistemic understanding determines one's evidential standards (Chinn et al., 2011), in other words, what one considers as evidence. Therefore, epistemic understanding is fundamental in argument skill, affecting the

strategies that an individual employs (Mason & Scirica, 2006), how one supports one's own position, and what kind of evidence one finds convincing.

Previous research has showed that epistemic understanding of what constitutes acceptable claims to knowledge, and ways to advance them in discourse, develop when individuals have the opportunity to engage in practice in dialogic argumentation and reflection while arguing, over an extended period of time (Kuhn et al., 2013). In the present work, the role of reflection beyond practise is examined, for its promotion of epistemic understanding.

Argumentation constitutes fertile ground to question beliefs, both one's own and others' (Grant, 2021). For example, when someone you disagree with tries to persuade you by providing their evidence, you have to examine the relation between the evidence and their claim. For example, "Does this piece of evidence really support this position?", and "Can causal conclusions be drawn here?". In addition to focusing on the relation between evidence and claim in others' arguments, one may focus on the evidence itself: "Does this piece of evidence come from a trustworthy source?"; "Is it strong evidence?". Similarly, when the interlocutor asks us to present our evidence, or challenges either the quality of the evidence we present or the connection we are making between evidence and our claims, we have to reflect on and question our own beliefs.

We hypothesize that engagement in reflection during argumentation, that is, putting argumentation itself as an object of examination, amplifies the gains of argumentation. Without the time pressure to respond or the strong emotions that usually accompany the argumentation process, one has the opportunity to pay closer attention to others' arguments, how one has responded to them, and how others have responded to her/him. Focusing on the exchange of arguments and the critique offered, where a piece of information may receive different, or even contradicting, interpretations, is a facilitative condition to develop an appreciation of the social nature of knowledge. This appreciation involves an epistemic understanding that evidence is amenable to evaluation, and argumentation is the means to evaluate knowledge claims. Therefore, reflection on argumentation can support the development of a metacognitive disposition towards argumentation, that is, of an epistemic understanding of the point of argument, that involves appreciating argumentation as an activity which is worthy of the mental effort expended to engage in it. Although disposition cannot be realized without competence, disposition is equally or even more important than metacognition as competence, and requires specific attention (Barzilai & Ka'adan, 2017; Bråten et al., 2011; Chinn et al., 2020; Iordanou et al., 2020; Kuhn, 2020; Mason et al., 2010; Moshman, 2020).

## The present study

The present work examined the role of reflection on individuals' argument skill. Students engaged in reflection of their own dialog with the help of reflection sheets. This prompted consideration of whether they had addressed and weakened their opponents' positions and whether they did so by using evidence. Students worked with a collaborating same-side partner when engaged in reflective activities to promote externalization of their thinking. These features have been hypothesized to facilitate metacognition of argumentation, by making the critical features of competent argument skill open to deliberation. Building on theories proposing that both meta-strategic and epistemic metacognition supports performance at the strategic level (Kuhn, 2000), it was examined whether enhancing meta-level awareness through engagement in reflective practice, would support the development of

individuals' argument skill at the performance level, that is, employment of counterargument strategy and use of evidence.

Based on theories suggesting a prominent role for reflection in both meta-level awareness and cognitive development (Inhelder & Piaget, 1958; Veenman, 2017), on Vygotsky's theory of internalization from the social plane, and on empirical evidence from previous studies (Felton, 2004; Iordanou & Constantinou, 2015; Kuhn et al., 2008; Shi, 2019), we examined the role of engagement in reflection at the social level for supporting argument skill. We hypothesized that engagement in reflection with a collaborative partner would support the development of argument skill, by heightening participants' metacognition, as both competence and disposition, in argumentation (Kuhn et al., 2008, 2013; Ryu & Sandoval, 2012). Participants' ability to pay careful attention to their opponents' position and seek to challenge it, that Walton (1989) identified as the goal of skilled argumentation, was used as the major criterion to assess argument skill and its development. The key research question which the experimental design addresses is: Does individuals' argument skill benefit from engagement in reflection during dialogic argumentation? To examine this question, two same-age equivalent groups of young adolescents engaged in the "Argue with me" pedagogical method for the same amount of time. Students in the experimental group (A&R) engaged in reflective activities while engaging in dialogic argumentation, while students in the control group (A) engaged mainly in dialogic argumentation, without engaging in reflective activities. Participants in both conditions discussed the same topic, received the same instructions during argumentation, had the same pieces of information available and had equivalent amounts of time to engage in the intervention. In both conditions, participants were told their dialogs were in preparation for a final class-level debate, ensuring equivalence in terms of motivation in engaging in the intervention. For the discourse topic, a contemporary social topic was employed, focusing on the issue of immigration, and how a nation should decide on which people from other countries should be allowed to come to live in their country.

In the present work we also included an assessment of students' topic knowledge. Previous work, both theoretical (Chinn & Duncan, 2018) and empirical (Baytelman et al., 2020; Means & Voss, 1996), suggests that topic-specific prior knowledge is tied to reasoning on that topic. In addition, a number of researchers view engagement in argumentation as a way to promote knowledge gains (Andriessen & Baker, 2014; Asterhan & Schwarz, 2016; Bereiter & Scardamalia, 2018; Nussbaum & Sinatra, 2003; Reznitskaya & Gregory, 2013; Weinberger & Fischer, 2006). According to Walton (2000), there can be dialectical shifts from one type of dialog, such as persuasion (aiming to persuade the other party), to another type of dialog, such as information-seeking (aiming to acquire or give information), in the context of a single argumentation exchange. Iordanou et al. (2019a), in particular, showed gains in both argument skill and knowledge acquisition within a single activity, establishing that it is possible to accomplish both knowledge and skill goals in the context of a single curriculum.

Based on previous work showing argument skill and knowledge gains through the same intervention (Iordanou & Kuhn, 2020; Iordanou et al., 2019a; Larrain et al., 2021), in the present work the effect of the intervention with respect to both was assessed. We wished to unpack the findings of Iordanou et al. (2019a), that reported knowledge gains after participants engaged in an argumentative-based activity, which involved both dialogic argumentation and reflective activities, by examining the unique contribution of engagement in reflective activities in promoting knowledge gains. There are several possible explanations of why and how reflection on argumentation might facilitate knowledge gains. The first possible explanation is that engagement in reflective activities prompts participants to reflect



on how they responded to arguments opposing their position, making more salient the differences between opinions. Asterhan and Schwarz (2016) see the articulation of one's ideas, the act of providing reasons to other persons to show why their idea is faulty and to support one's own position after receiving critique from others, that takes place during argumentation, as facilitative conditions for learning. As students seek to reconcile the differences that are highlighted in the context of discussion, conceptual change starts to occur. A second possible explanation is that engagement in reflective activities involves writing – for registering one's arguments, others' responses and constructing a revised rebuttal – that Mason and Boscolo (2000) found facilitates conceptual understanding, which offers another advantage of engagement in reflective activities to promote knowledge gains compared to engagement in sole discussion. A third possible explanation, supported by self-regulation theories (e.g. Efklides, 2011; Zimmerman, 2000) is that reflection supports metacognitive awareness of lack of knowledge on the topic, which then motivates individuals to search for more information.

A fourth possibility is related to metacognition as disposition. When individuals are asked to explain their positions during engagement in reflective activities, either in the context of collaborating with a same-side partner or when trying to revise their response after being invited by the interlocutor, they may become less certain of their view, according to the illusion of explanatory depth (Rozenblit & Keil, 2002). They may also become more willing to examine alternative views or search for more evidence to be able to address the “how” questions of the interlocutor; “How do you know that?”. A fifth possibility, which is relevant to the fourth, is that being prompted to reflect on one's own and others' arguments may support rethinking of one's own position, in the light of opponents' critiques of our claim, for example questioning the quality of the evidence we provided, or observing that the interlocutor offered an alternative interpretive framework of a particular set of data that seems more powerful than the one we have provided. This realization that reality is complex and there are alternative interpretations, some of which might be more powerful than one's own, may lead to an appreciation of what Grant (2021) called “humility about our knowledge” (p. 158), creating more doubts about our opinions and triggering our curiosity, which will motivate us to seek information we are lacking. Therefore, reflection might support the development of an evaluativist epistemic understanding, that involves a mindset of searching for more evidence and considering the power of different interpretative frameworks before making a judgment; a mindset which encourages seeking further knowledge and contributing to knowledge gains and conceptual gains. Therefore, we hypothesize that engagement in reflective activities will contribute above and beyond engagement in discussion to knowledge acquisition. Given our limited understanding of *how* engagement in argumentation promotes knowledge acquisition (Asterhan & Schwarz, 2016; Van der Veen & Van Oers, 2017; Wecker & Fischer, 2014), the present study has a unique contribution to offer in the literature by examining the role of reflection on argumentation for promoting knowledge gains.

This is the first time we assessed content knowledge, after engagement in the “Argue with me” curriculum. Previous work assessed content knowledge by examining participants' arguments (Iordanou et al., 2019a), rather than employing a knowledge test. Knowledge gains were assessed using a multiple-choice test which directly examined content knowledge about the intervention topic. Gains in argument skill were examined by applying a coding scheme to the written arguments that individuals produced before and after their engagement in the intervention. The coding scheme has been used extensively in previous research to examine gains in argument skill as a result of engagement in interventions similar to the one employed here (Hemberger et al., 2017; Kuhn et al., 2016; Iordanou



et al., 2019a; Shi et al., 2019), especially in respect of addressing opposing arguments and seeking to weaken them by using evidence. To examine the process of development of both argument skill and meta-level awareness, the microgenetic method was employed to examine participants' dialogs over the course of the intervention. This enabled direct observation of the process of change as individuals engage in the same task repeatedly (Kuhn, 2000). Participants were asked to conduct their discourse electronically, a technique which has the benefit of providing an immediately available and permanent record of the discourse that participants can use as the basis of their reflective activities. It also facilitates the microgenetic examination of the dialog during the intervention.

## Method

### Participants

Participants were 41 elementary school students (ages 11–12; 21 males), from two equivalent 6<sup>th</sup> grade classes (of 21 and 20 students) from 2 schools in Cyprus. The schools were non-selective public schools, serving middle-class families. Participants' performance on the initial essay supported the initial equivalence of groups (see [Preliminary analysis](#)). Participants were permanent residents of Cyprus, born and grown up in Cyprus, with primarily an average academic achievement, typical of these schools. Participants had attended the same school and had been part of the same class since grade one. Participants were recruited for the purposes of the present study, after securing ethical approval from the Cyprus National Ethics Committee and Ministry of Education. All the students in each class provided written parental consent to participate in the study. One of the participating classes was randomly assigned to serve the experimental condition (A&R), while the other class served in the control condition (A).

### Initial assessment

#### Assessing individual argument skill

To assess participants' argument skill, participants were asked to write a brief essay taking a position on the intervention topic, namely how a nation should decide which people from other countries should be allowed to come to live in their country. The question was whether a nation should allow immigrants based on what they can contribute to the nation, or how bad life is where they come from. Participants were introduced to the topic by reading two short passages on the topic of immigration; one discussed cases of countries which accepted immigrants based on the receiving country's needs and what those people could offer to the recipient country, while the other discussed cases of countries that welcomed immigrants based on how bad the situation in their country was, due for example to financial issues. Participants were asked to take a position and write the argument they would make to someone who didn't agree that their position was the better one.

#### Assessing prior knowledge

Prior to assessing participants' individual argument skill, their topic knowledge was assessed. To assess participants' topic knowledge, participants were asked to answer a

multiple-choice test that was developed for the needs of the present study, including 10 questions on the topic of immigration. The knowledge test was administered both before the intervention, to assess participants' prior knowledge on the topic, and after the intervention, to assess their learning. The knowledge test items addressed information that was provided to participants on cards during the intervention, testing recall of a piece of information, and conceptual understanding, which involve processing and integration of multiple pieces of information that were available to participants (e.g., what immigrants can contribute to the welcoming country?). Participants received one point for each correct answer. The reliability (Kuder-Richardson 20) for scores on the prior knowledge test was 0.62. While this reliability estimate is somewhat lower than desirable, it can be considered acceptable for research purposes (Hair et al., 1998; Kerlinger & Lee, 2000).

## **Intervention**

The intervention took place over thirteen twice-weekly class periods, each of 80-min. Each condition (A&R and A) worked independently of the other condition. These began with Session 1 in the week after to the initial assessment sessions. All participants were told that the purpose of the activity was to learn about the topic and to prepare for a whole-class verbal debate, where their parents would be invited to attend. Participants in each condition/class (A&R or A) were assigned to one of two teams based on the participants' position, as reported in the initial essay; the welcoming country's needs (WCN), or bad conditions in the immigrants' country (BC). The classroom teacher and a research assistant facilitated the intervention as adult coaches, answering any questions and reminding pairs to collaborate with one another. The coaches were blind to the hypothesis of the study, and were provided a detailed intervention protocol, including lesson plans and assessment guidelines. Meetings were held with a senior researcher before and after each session.

### **Preparation: session one**

Participants in both conditions/classes (A&R or A) assembled randomly into same-side groups of 4–5 participants (WCN or BC), received some background information about the topic and were asked to generate reasons which supported their side's preferred position. Participants were asked to record their reasons on cards, eliminate duplicates and rank their reasons with respect to their strength.

### **Dialogs and reflection: sessions two to nine**

Each same-side group (WCN / BC) was divided into pairs who worked together throughout the intervention. In the experimental condition (A&R, 21 students) we had 10 teams – 9 pairs and one group consisting of 3 students (which we continue to refer to as a pair for reporting purposes); in the control condition (A, 20 students) we had 10 pairs. Each same-side pair (WCN / BC) engaged in an electronic dialog via an instant-messaging platform, using tablets, with a sequence of five opposing-side pairs (WCN / BC). Over eight 80-min sessions, the pair argued with an opposing-side pair that was different to the one they argued with in the previous session, arguing once or twice with each of the five opposing-side pairs over the course of the 8 sessions. In rare cases that one pair member was absent, the other member engaged individually in the discussion. Participants were asked to work collaboratively with their partner in deciding what to

communicate to the other pair they were engaged in dialog with, and once in agreement to type their response. They were instructed to try to convince the other pair that their position was the better one. Dialogs lasted approximately 35 minutes. Participants had available some relevant information in the form of a set of cards, which was balanced overall with respect to support for the two positions. Each card contained a question and a short answer (e.g. “What kinds of jobs do immigrants do in Cyprus?”, and “Can anyone who wishes decide to immigrate to Cyprus?”). Some of the questions were based on questions in Kuhn (2018), while others were developed for the purpose of the present study using reliable sources, with source information provided under each answer. At each session, the cards that were provided in previous sessions remained available to participants and three new cards were provided to them. Participants were encouraged to read the cards and use them if they wished, but no explicit instructions were provided with respect to argumentation.

During Sessions 5–9, participants in the experimental condition/class (A&R) engaged in some reflective activities alongside the discussions, while participants in the control condition/class (A) engaged only in discussions. While waiting for the other pair to respond, A&R experimental condition participants were asked to reflect on the electronic transcript of their own dialog, which was available to them in the instant-messaging platform on their tablets. To support engagement in reflection, participants were asked to complete one of two reflection sheets (Jordanou & Constantinou, 2015), which alternated across sessions. Own-side reflection sheets asked participants to reflect on (a) whether they had paid attention and addressed the opposing side’s position, and (b) whether they had used *evidence* in the counterarguments they had constructed to the opposing side’s position, focusing on an excerpt of their dialog that they had chosen. Other-side reflection sheets asked participants to reflect on the rebuttals they used to weaken counterarguments to their own position, and on whether they used *evidence* to support their rebuttal. In each case, they were asked to contemplate what a better counterargument or rebuttal might have been.

### Preparation for showdown: session 10

In Session 10, same-side groups reassembled and worked together to prepare for the whole-class debate. A&R experimental group participants had available the reflection sheets and printed transcripts of the dialogs they completed in previous sessions. Control group participants (A) had available printed transcripts of their dialogs. The group had to decide what arguments to use in the debate. Colored cards were made available to participants, and they were encouraged to use different colored cards to summarize arguments, counterarguments and rebuttals, as well as evidence supporting their claims. An adult coach (a research assistant) facilitated these discussions.

### Showdown and feedback: sessions 11 to 13

The week after Session 10, in Session 11, the electronic Showdown, participants engaged in a class-level debate on the computer – messaging –, between participants holding opposing positions who were sited in different rooms. In each room, half of the participants participated in debate for the first 45 minutes of the session, while the other half of the participants engaged in the debate the last 45 minutes. One member of the group was designated as the typist. In the session following the electronic Showdown, Session 12, an argument map was presented to the participants, prepared by the researchers. Different colors

were used to indicate effective and less effective argumentive moves as well as whether evidence was used effectively to support participants' arguments and critique. During the final session of the intervention, Session 13, participants had a face-to-face debate, which was attended by the participants' parents.

### Final assessment

Final assessment was identical to initial assessment. Participants in both conditions A&R and A, were asked to write a final essay on the discourse topic. The prompt was the same as the initial assessment; to "Write the argument you would make to someone who didn't agree that your position is the better one." Participants repeated the multiple-choice test assessing their topic knowledge.

### Coding of essays at initial and final assessment

Each initial and final essay was divided into idea units, with an idea unit defined as an assertion with any accompanying justification. The author and another coder, blind to condition and time, segmented the essays of the participants into idea units. Interrater reliability on segmenting was achieved on a subset of 30% of the essays with 90% agreement (Cohen's Kappa = 0.88). One of the coders proceeded to segment the remaining essays, blind to condition and time. Idea units were further coded on (a) whether they included evidence, (b) whether the evidence was functional, and (c) the function that the evidence served. Interrater reliability on coding was achieved, with 87% agreement (Cohen's Kappa = 0.84), on a subset of 30% of units. One of the coders proceeded with coding the remaining essays, again blind to condition and time. The evidence that participants employed came mostly from the information cards that were provided, and from their personal knowledge. If a piece of information was cited but was not connected to any claim, this was coded as a non-functional unit and not analysed further. The units which contained a claim and accompanying evidence used in the service of the claim, were classified as functional units. Functional units were further classified as serving one of four mutually exclusive functions, using a coding scheme adapted from that reported in previous research (Iordanou et al., 2019a; Kuhn et al., 2016). Evidence functioning to support the student's own chosen position on the topic was coded as Support-Own; evidence functioning to weaken the opposing position on the topic was coded as Weakening-Other; evidence functioning to support the opposing position on the topic was coded as Support-Other, and evidence functioning to weaken the student's own position on the topic was coded as Weakening-Own. The coding scheme is based on the rationale that skilled argument requires attention to all four argument functions.

### Coding of electronic discourse during the intervention

All the online dialogs that took place between two pairs holding opposing positions during the intervention were segmented into idea units and coded. Each idea unit was classified as to whether or not it constituted metatalk, defined as statements *about* the dialog, in contrast to dialog statements about the topic content itself (Kuhn et al., 2008). Each metatalk utterance was further categorized as either meta-strategic or

epistemic. Meta-strategic utterances seek to monitor or direct the discourse performance (e.g. “You haven’t responded to my argument”). Epistemic statements seek to judge the type and quality of the information provided to be used as evidence for supporting a claim (e.g. “That’s your opinion, we want statistics”; “Did you find this in a particular source or you made it up?”). Dialog statements were coded using the same coding scheme used to code participants’ initial and final assessment essays: (a) whether they included evidence, (b) the evidence was functional, and (c) the function that the evidence served (Support-Own, Weakening-Other, Support-Other and Weakening-Own). The coders coded the electronic discourse blind to time and condition. Interrater agreement, based on 50% of the data, was 95% (Cohen’s Kappa = 0.89). The rest of the data were coded by one of the coders. Disagreements were resolved through discussion.

## Results

An examination of participants’ initial and final individual essays and multiple-choice knowledge tests served as the main indicators to assess participants’ argument skill and knowledge acquisition, respectively, and for comparison of differences across conditions.

Our initial analytic task was to apply a common coding scheme to the initial and final essays, which allows us to examine overall progress in argument skill, before micro-genetic examination could begin. Two students from the experimental condition (A&R) and one student from the control condition (A), who were receiving special education, requested to participate in the study but not in the initial and final assessment, therefore are not included in the analysis of the study. The analysis therefore included 19 students in each condition.

### Knowledge acquisition

To examine if participants gained content knowledge on the intervention topic of immigration, and if the condition had an effect on participants’ knowledge gains, a repeated-measures ANOVA was used, using participants’ score on the multiple-choice test completed at initial and final assessment as the dependent variable. ANOVA results showed a significant Time X Condition interaction,  $F(1, 36) = 4.167$ ,  $p = 0.049$ ,  $\eta_p^2 = 0.104$  and a main effect of Time,  $F(1, 36) = 52.807$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.595$ .<sup>1</sup> At initial assessment, participants in both conditions exhibited similar performance, getting one third of their responses right,  $M = 3.316$  ( $SD = 1.157$ ) for the experimental condition (A&R), and  $M = 3.47$  ( $SD = 1.54$ ) for the control condition (A). At the final assessment, all participants showed improvement, with participants in the experimental condition (A&R) outperforming participants in the control condition (A),  $M = 6.316$  ( $SD = 1.183$ ) and  $M = 5.158$  ( $SD = 2.141$ ), respectively.

<sup>1</sup> Checking the internal homogeneity of the clusters, the ICC for content knowledge was .19, substantially above the recommended value of .05 (Thomas & Heck, 2001), therefore the results should be interpreted with caution.

## Argument skill

### Preliminary analysis

Data were examined for outliers, both univariate and multivariate, and these were ruled out. We first compared the two groups' performance at initial assessment to ensure that participants in the two conditions performed equivalently. Independent sample t-tests showed that there was no significant difference between experimental (engaged in Argumentation & Reflection) and control (engaged only in Argumentation) groups in the number of coded units,  $t(36)=1.407$ ,  $p=0.168$ , the number of evidence-based units,  $t(36)=-0.798$ ,  $p=0.430$ , and the number of functional units,  $t(36)=-0.383$ ,  $p=0.704$ . No significant difference between the experimental (A&R) and control (A) conditions was observed at initial assessment in the units which functioned to support own position,  $t(36)=-1.302$ ,  $p=0.201$  and weaken other position,  $U=179.50$ ,  $p=0.956$ . No differences were observed either in the usage of units functioning to Support-Other position and Weakening-Own position, for which none of the participants in either the experimental (A&R) or control (A) condition used these types of units.

### Number of units

To assess the effects of group and time on the mean number of units produced in participants' essays, a generalized linear mixed model (GLMM), using the Poisson probability distribution, was used, with the individual as the unit of analysis. A GLMM showed that the overall model was significant,  $F(3, 72)=12.830$ ,  $p<0.001$ . The interaction between group and time was not statistically significant,  $F(1, 72)=2.293$ ,  $p=0.134$ . The fixed effect of group was not statistically significant,  $F(1, 72)=0.000$ ,  $p=1.000$ , while the fixed effect of time was statistically significant,  $F(1, 72)=34.673$ ,  $p<0.001$ . The mean number of units in the experimental condition (A&R) participants' essays increased from  $M=1.526$  ( $SD=0.213$ ), 95% CIs [1.101, 1.952], to  $M=3.842$  ( $SD=0.391$ ), 95% CIs [3.064, 4.621]. For essays of participants in the control condition (A) this mean increased from  $M=2.000$  ( $SD=0.244$ ), 95% CIs [1.513, 2.487] to  $M=3.368$ , ( $SD=0.366$ ), 95% CIs [2.639, 4.097].

### Evidence-based units

Regarding evidence-based units, GLMM, using the Poisson probability distribution, showed that the overall model was significant  $F(3, 72)=19.576$ ,  $p<0.001$ . The time condition interaction was not statistically significant,  $F(1, 72)=0.740$ ,  $p=0.393$ . The main effect of time was statistically significant  $F(1, 72)=56.990$ ,  $p<0.001$ , while the main effect of condition was not statistically significant,  $F(1, 72)=2.639$ ,  $p=0.109$ . Participants in both the experimental (A&R) and control (A) conditions showed improvement from initial to final assessment, from  $M=1.421$  ( $SD=0.198$ ), 95%, CIs [1.027, 1.815] to  $M=3.737$  ( $SD=0.356$ ), 95%, CIs [3.026, 4.447] and from  $M=1.211$  ( $SD=0.183$ ), 95%, CIs [0.847, 1.575] to  $M=3.053$  ( $SD=0.322$ ), 95%, CIs [2.410, 3.695], respectively.

## Functional units

Regarding frequencies of functional usage, a GLMM using the Poisson distribution, showed that the overall model was significant,  $F(3, 72)=17.669$ ,  $p<0.001$ . The interaction between condition and time was not significant,  $F(1, 72)=2.843$ ,  $p=0.096$ . The main effect of time was significant,  $F(1, 72)=51.679$ ,  $p<0.001$ , while the main effect of condition was not  $F(1, 72)=4.167$ ,  $p=0.045$ . Experimental condition participants (A&R) improved from  $M=1.105$  ( $SD=0.200$ ), 95% CIs [0.706, 1.505], to  $M=3.737$  ( $SD=0.402$ ), 95% CIs [2.936, 4.538]. Control condition (A) participants improved from  $M=1.000$  ( $SD=0.191$ ), 95% CIs [0.620, 1.380], to  $M=2.632$  ( $SD=0.337$ ), 95% CIs [1.959, 3.304].

In sum, both groups exhibited an increase in the number of overall idea units produced, the number of evidence-based units and functional units produced, from initial to final assessment. Below we examined whether there were any condition differences in the function of those evidence-based claims.

## Support-own usage

Starting from the most common and less challenging function, namely the use of evidence to support one's own condition, a GLMM using the Poisson probability distribution for number of units which functioned to Support-Own condition showed that the model was significant,  $F(3, 72)=6.189$ ,  $p=0.001$ . The time condition interaction was not significant,  $F(1, 72)=0.217$ ,  $p=0.643$ . The main effect of time was significant,  $F(1, 72)=15.669$ ,  $p<0.001$ , while the main effect of condition was not significant,  $F(1, 72)=0.868$ ,  $p=0.355$ . Experimental condition (A&R) participants improved from  $M=1.000$  ( $SD=0.190$ ), 95% CIs [0.621, 1.379] to  $M=1.789$  ( $SD=0.272$ ), 95% CIs [1.248, 2.331]. Similarly, control condition (A) participants improved from  $M=0.684$  ( $SD=0.157$ ), 95% CIs [0.370, 0.998] to  $M=1.684$  ( $SD=0.264$ ), 95% CIs [1.159, 2.210]. Results showed comparable performance in the two conditions, regarding the less demanding function of using evidence to support one's own position.

## Weakening-other usage

Are there any differences between the two conditions regarding the most challenging function which requires attention to the other's position? A GLMM provided an affirmative answer, showing that there was a significant difference between the two conditions in Weakening-Other usage,  $F(3, 72)=9.726$ ,  $p<0.001$ . The interaction between condition and time was significant,  $F(1, 72)=7.095$ ,  $p=0.010$ . Both the fixed effect of time,  $F(1, 72)=28.381$ ,  $p<0.001$ , and the fixed effect of condition were significant,  $F(1, 72)=5.864$ ,  $p=0.018$ . Although participants in both conditions showed a significant increase from initial to final assessment, participants in the experimental condition (A&R) showed greater improvement, from  $M=0.105$  ( $SD=0.084$ ), 95% CIs [-0.062, 0.272] to  $M=1.842$  ( $SD=0.350$ ), 95% CIs [1.145, 2.540], compared to participants in the control condition (A), from  $M=0.158$  ( $SD=0.103$ ), 95% CIs [-0.047, 0.363] to  $M=0.737$  ( $SD=0.221$ ) 95% CIs [0.296, 1.178].



## Support-other and weakening-own usage

The usage of units which functioned to either Support-Other position or Weakening-Own position was negligible in both the experimental (A&R) and control (A) conditions, with no usage at initial assessment and very limited usage at the final assessment (see Table 1).

## Individual patterns of change

In addition to analyses of group trends, analyses of changes at the individual level are equally informative. Table 2 shows the number of participants who used the different types of functional units at least once. As seen in Table 2, although there was no difference in the number of participants who used Weakening-Other units at initial assessment,  $n=2$  for both conditions, there was a significant difference at the final assessment, where more experimental condition (A&R) participants ( $n=14$ ) used Weakening-Other units compared to control condition (A) participants ( $n=9$ ), ( $p=0.033$ , Fisher's Exact Test). Those results confirmed the group-level results.

## Microgenetic analysis: argument skill development during the intervention

The microgenetic method was employed to examine the process of change during the intervention. Note that quantitative analysis using the individual as the unit of analysis was not

**Table 1** Mean number of Idea Units, Evidence-Based Units, Functional Units, Support-Own, Weakening-Other, Weakening-Own and Support-Other Evidence Units in Initial and Final Essays

	Experimental Condition (A&R)		Control Condition (A)	
	Initial Assess	Final Assess	Initial Assess	Final Assess
Idea Units	1.526 (.213)	3.842 (.391)	2 (.244)	3.368 (.366)
Evidence-based Units	1.421 (.198)	3.737 (.356)	1.211 (.183)	3.053 (.322)
Functional Units	1.105 (.200)	3.737 (.402)	1 (.191)	2.632 (.337)
Support-Own	1 (.190)	1.789 (.272)	.684 (.157)	1.684 (.264)
Weakening-Other	.105 (.084)	1.842 (.350)	.158 (.103)	.737 (.221)
Support-Other	0	.105 (.459)	0	.211 (.918)
Weakening-Own	0	0	0	.053 (.229)

**Table 2** Number of Participants who Used Evidence-based Units, Functional Units, Support-Own, Weakening-Other, Weakening-Own and Support-Other Evidence Units at least once

	Experimental Condition (A&R)		Control Condition (A)	
	Initial Assess	Final Assess	Initial Assess	Final Assess
Evidence-based Units	18	19	13	19
Functional Units	15	19	12	18
Support-Own	14	16	10	18
Weakening-Other	2	14	2	9
Support-Other	0	1	0	1
Weakening-Own	0	0	0	1

possible because participants were working in pairs, which remained fixed throughout the intervention; therefore, the pair was used as the unit of analysis.

We present experimental (A&R) and control (A) group performance for each of the eight dialog sessions of the intervention to examine if there is any pattern of change. Table 3 shows participants' use of evidence-based units, functional units, and the different types of functional units (Support-Own, Weakening-Other, Support-Other, and Weakening-Own). As seen in Table 3, the two conditions showed comparable performance in the use of evidence-based units, functional units and evidence-based units, which served to support their own positions throughout the intervention. However, a different pattern was observed in the usage of the Weakening-Other strategy between the two conditions. Figure 1 shows the usage of the Weakening-Other strategy throughout the intervention, including also initial and final assessment. As seen in Fig. 1, experimental (A&R) and control (A) condition participants exhibited comparable patterns in the usage of evidence to weaken other's arguments during the first four dialog sessions, while a difference is observed in the last 4 dialog sessions of the intervention, with experimental condition participants (A&R) showing greater usage compared to control (A) condition participants.

The greater improvement exhibited by experimental condition participants (A&R) compared to control (A) condition participants in the last 4 dialog sessions (Dialog sessions 5 to 8) can be interpreted as due to the reflective activities that experimental condition participants but not control condition participants engaged in from the 4<sup>th</sup> to the 8<sup>th</sup> dialog session.

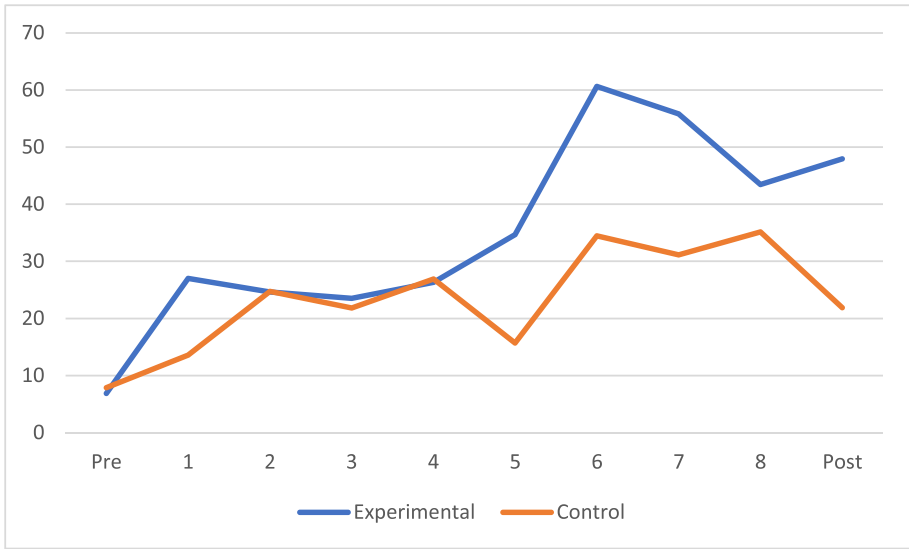
### Meta-level development during the intervention

To examine further the hypothesis of the present study and the question of mechanism identified as central to the present work, in this section of the analysis our aim was to examine the extent to which participants showed progress at the meta-level – reflected in the usage of meta-level talk – as distinguished from progress in implementing successful argumentative strategies, at the strategic level. We expected progress of the former type to be most visible in the experimental condition (A&R).

As seen in Figs. 2 and 3, which show the pattern of overall usage of epistemic and meta-strategic talk per session, respectively, during the first four dialog sessions (1–4) experimental (A&R) and control (A) condition participants showed a comparable pattern of usage of meta-level discourse. Reflective activities were introduced in the experimental condition only (A&R) in Dialog Session 4, and from this point the two groups' pattern of usage of meta-level talk differed. As was the case in previous usage of the microgenetic method to examine argument skill (see Kuhn et al., 2008) variability is the norm, rather than consistent upward change. Using the pair as the unit of analysis, experimental condition (A&R) students' meta-strategic usage increased from  $M=0.625$  ( $SD=1.976$ ) at initial assessment to  $M=5.929$  ( $SD=9.875$ ) at the final assessment, while control (A) condition students showed no change, from  $M=2.361$  ( $SD=4.988$ ) to  $M=2.778$  ( $SD=6.000$ ), respectively. The same pattern was observed for epistemic usage, with experimental condition students (A&R) showing an increase from  $M=0$  to  $M=8.571$  ( $SD=18.07$ ), while control condition students showed similar performance at initial ( $M=3.25$ ;  $SD=7.076$ ) and final assessment ( $M=2.778$ ;  $SD=6.000$ ). A repeated measures ANOVA on meta-level talk – the sum of meta-strategic and epistemic talk – showed a significant Time X Condition interaction,  $F(1, 18)=4.715$ ,  $p=0.044$ ,  $\eta_p=0.208$ . Experimental condition (A&R) participants exhibited a significant increase from initial ( $M=0.625$ ,  $SD=1.976$ ) to final

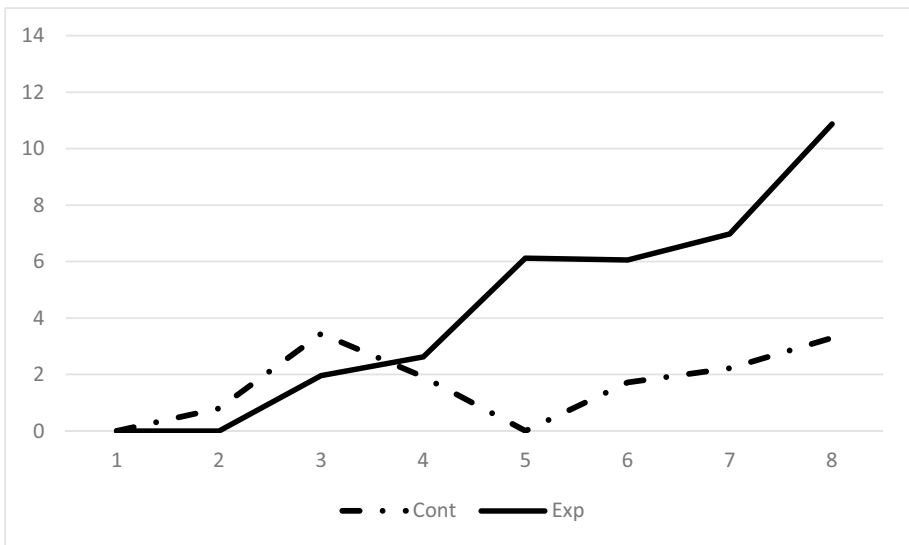
**Table 3** Percentage (and Frequency) of Evidence-based Units, Functional Units, Support-Own, Weakening-Other, Support-Other and Weakening-Own Units over Time

	Dialog 1		Dialog 2		Dialog 3		Dialog 4		Dialog 5		Dialog 6		Dialog 7		Dialog 8	
	Exp	Con	Exp	Con	Exp	Con	Exp	Con	Exp	Con	Exp	Con	Exp	Con	Exp	Con
Idea Units	74	81	65	93	51	87	38	52	49	51	33	39	41	45	46	91
Evidence-based Units	57%(42)	37%(30)	52%(34)	53%(49)	35%(18)	49%(43)	47%(18)	60%(18)	61%(30)	55%(28)	79%(26)	64%(31)	71%(29)	53%(24)	65%(30)	59%(54)
Functional Units	57%(42)	35%(28)	49%(32)	53%(49)	35%(18)	47%(41)	47%(18)	60%(31)	59%(29)	51%(26)	76%(25)	64%(31)	71%(29)	53%(24)	65%(30)	57%(52)
Support-Own	30%(22)	21%(17)	23%(15)	25%(23)	22%(11)	25%(22)	21%(8)	33%(17)	18%(9)	35%(18)	15%(5)	31%(18)	15%(6)	22%(10)	17%(8)	22%(20)
Weakening-Other	27%(20)	14%(11)	25%(16)	25%(23)	24%(12)	22%(19)	26%(10)	27%(14)	35%(17)	16%(8)	61%(20)	33%(13)	56%(23)	31%(14)	44%(20)	35%(32)
Support-Other	1%(1)	0	2%(1)	3%(3)	0	1%(1)	3%(1)	0	6%(3)	0	0	0	0	2%(1)	4%(2)	2%(2)
Weakening-Own	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

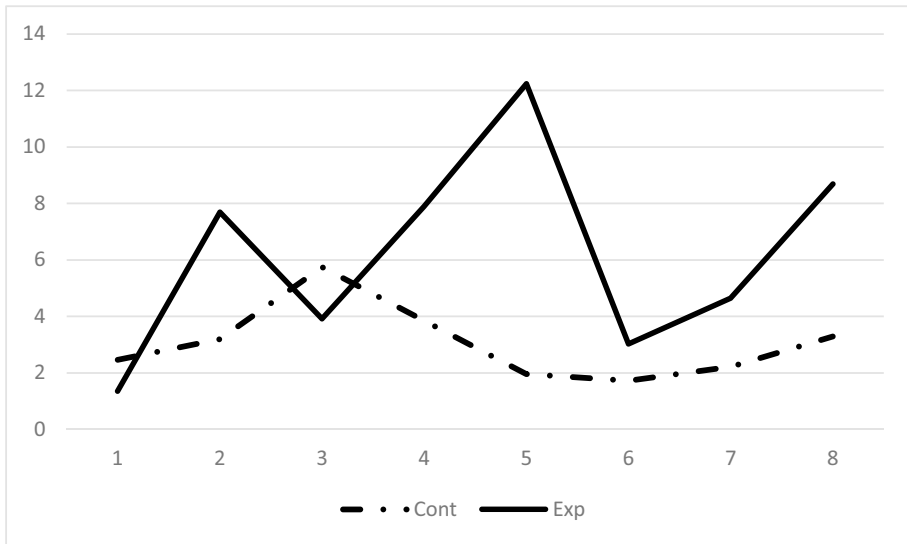


**Fig. 1** Percentage of Units That Functioned as Weakening-Other Throughout the 8 Dialog Intervention Sessions and at Initial and Final Assessments

assessment ( $M = 14.500$ ,  $SD = 17.639$ ), whereas no significant change was observed in control (A) condition participants (from  $M = 5.611$ ,  $SD = 7.609$  to  $M = 4.722$ ,  $SD = 6.425$ ). The components of meta-talk (epistemic and meta-strategic) did not show a statistically significant change. Interestingly, during Dialog Sessions 4 to 8 of the intervention, experimental participants (A&R) showed greater usage of meta-level talk, as seen in Figs. 2 and 3, but



**Fig. 2** Percentage of Overall Units Coded as Epistemic Discourse in Experimental and Control Conditions, Over Time (Dialog Sessions) During the Intervention



**Fig. 3** Percentage of Overall Units Coded as Meta-Strategic in Experimental and Control Conditions, Over Time (Dialog Sessions) During the Intervention

also greater usage of the Weakening-Others argumentative strategy, as seen in Fig. 1, providing support to our hypothesis that meta-level understanding co-develops and supports development in argumentative strategies.

Below is an example of talk from experimental condition students' discussion in the final Dialog Session 8 of the intervention, which provides some insights into their meta-level development.

Pair A: Yes, but enough immigrants and refugees have quite good jobs such as skilled production workers (16.2%).

Pair B: But this is a very small percentage.

Pair A: There are many jobs they could do, this was just an example.

Pair B: Which ones?

Pair A: These: unskilled workers (38.6%), service and sales employees (18.6%), skilled production workers (16.2%), qualified jobs (7.6%).

Pair B: Yes, but these are old data.

Pair A: They have been published in 2016!

Pair B: Several years have passed (since then). In other sources it is mentioned that some immigrants went to other countries or have died since 2016, so the percentages are lower.

Initially, Pair A provided a percentage to serve as evidence to support their point. Then Pair B critiqued the evidence that Pair A offered ("this is a very small percentage") and Pair A offered a rebuttal to pair B's critique ("this was just an example"). Then Pair B critiqued the credibility of the evidence, making a comment about the date of publication of this information "these are old data". Pair B continued this critique by highlighting that there are other more recent sources which include information that contradict the conclusions drawn if one relies only on Pair A's source. Students' epistemic talk in this example provides insight into students' epistemic standards. In particular, the students appear to

share the following epistemic standards: percentages are a good piece of evidence to use to support a claim; the date of the source of information should be taken into consideration, with recent sources being more credible than older ones; multiple sources should be taken into consideration, and conclusions should be drawn after integration of information from multiple sources.

## Discussion

The findings of the present work highlight the role of *reflection* in supporting the development of argument skill and meta-level awareness of argumentation. We start the discussion by examining the gains observed in the experimental condition (A&R) in comparison with the control (A) condition, which differed only with respect to engagement in *reflective activities* alongside engaging in dialogic argumentation. This comparison will reveal how engagement in reflection has affected argument skill. Then we examine changes which took place at both the procedural and meta-level during the intervention, in order to obtain some insight into the *mechanism* underlying the gains observed at the strategic level in argument skill from initial to final assessment. Finally, educational implications are discussed.

Did engagement in reflective activities affect participants' argument skills? Our findings provide an affirmative answer to this question. Despite equivalence of purpose, engagement in dialogic argumentation, available information to use as evidence, and instructions across conditions, the difference in engagement in reflection resulted in stronger argument skill on the intervention topic among participants who engaged in reflection in addition to argument. Experimental condition participants who engaged in reflective activities about argumentation alongside engaging in dialogic argumentation outperformed their counterparts who engaged only in dialogic argumentation. The use of evidence with the purpose of weakening others' positions, which lies at the heart of argument skill, increased in frequency more among experimental condition participants. The improvements at the individual level also favour the experimental condition. At initial assessment only 10% of participants in both conditions used evidence to counter others' positions. At the final assessment, the majority of experimental condition participants (75%) showed skilled usage of evidence to weaken others' positions, compared to only 50% of control condition participants. The increase in usage of evidence-based arguments, and particularly evidence weakening others' positions, is consistent with earlier studies involving similar methods (Arvidsson & Kuhn, 2021; Hemberger et al., 2017; Iordanou & Kuhn, 2020; Iordanou et al., 2019a; Murphy et al., 2018; Shi, 2019; Shi et al., 2019).

Gains were also observed in the control condition, which showed comparable improvements with the experimental condition in using evidence, both in functional use of evidence and for using evidence to support their own position. The improvements exhibited by the control group which engaged only in dialogic argumentation, shows that engagement in extensive practice in argumentation with peers holding an opposing view is beneficial for supporting individuals' argument skill, a finding which is consistent with other empirical findings in the literature (Felton, 2004; Fisher et al., 2017), and with Vygotsky's sociocultural theory (1978). However, experimental condition participants' gains show that a combination of engagement in dialogic argumentation practice *and* reflective activities is a more powerful method for promoting argument skill, compared to practice alone.

Our experimental design, which put reflection under rigorous investigation, reveals the unique contribution of reflection beyond other features – goal-based nature of

activity, motivational elements, collaborative work, practice in dialogic argumentation – in supporting argument skill development on the intervention topic. This study extends the line of research examining how argument skill develops, and particularly which elements are fundamental for supporting the development of argument skill. Another study, for example, revealed that engagement in discussions with peers holding an opposing view is more beneficial in developing argument skill compared to engagement with same-side peers (Iordanou & Kuhn, 2020). The method of providing information in the form of question-and-answer has been found superior to a traditional text-based method to promote acquisition of factual knowledge sufficient to support argumentation (Iordanou et al., 2019a). The present study reveals the unique role of engagement in reflective activities, beyond dialogic practice, in employing evidence to support critique of opponents' positions. A limitation of the present work is that it examined argument skill on the intervention topic, rather than on a new topic, as was the case in previous studies (Iordanou & Kuhn, 2020; Iordanou et al., 2019a). Although previous studies that employed the "Argue with me" curriculum showed transfer of gains in argument skill to a non-intervention topic (see Iordanou & Rapanta, 2021, for a review), the differences observed in the current study between experimental and control conditions, which differ only in respect to engagement in reflection, to a non-intervention topic remain to be explored. Some other limitations involve the small sample size and the nested structure (Cress, 2008) of the data (e.g. participants were nested in dyads, in groups of four, in different classes), that can increase the Type I error rate. These limitations do not enable us to rule out possible environmental differences (e.g. teacher effects) and selection-history threat to internal validity. Further, given that the reflective activities employed prompts to engage participants in reflection, future research could investigate the contribution of prompts in both argument skill and knowledge gains (Iordanou et al., 2019a), in addition to or in combination with engagement in reflection.

The novelty of the present study lies in the examination of discourse during the intervention. We now move to identify any pattern which might have helped in interpreting the condition differences observed in post-discourse individual argument skill. Using the microgenetic method, our data showed a different pattern of progress between the two conditions when the reflective activities were introduced in the experimental condition, with experimental condition participants showing greater improvements at both the strategic and meta-strategic level. In particular, when engaged in reflective activities, during Dialog Sessions 4–8 of the intervention, experimental condition participants showed greater usage of both meta-level talk and the advanced argumentative strategy of employing evidence to weaken others' positions. These data support the hypothesis of the present study that engagement in reflection while individuals are engaged in dialogic argumentation fosters the development of an epistemic disposition to use and critique evidence and the norms of argumentation, which supports performance at the strategic level.

The patterns of development of meta-level understanding during the intervention inform our understanding of *metacognitive development* and how it can be supported. Engagement in tasks which involve the targeted strategies supports the development of metacognitive abilities, as the control condition participants' performance shows; this finding is consistent with other findings in the literature (Schneider & Pressley, 1997; Tsalas et al., 2017), and supports Flavell's (1979) conception of metacognition. Besides practice in argumentation, we cannot rule out the possibility that other features of the curriculum may have supported individuals' meta-level awareness of argumentation, such as the presence of a same-side peer with whom they collaborated, discussion with peers who hold an opposing position



(Jordanou & Kuhn, 2020), which might have provided feedback on whether their argumentative moves were persuasive or not, and that the communication took place through instant messaging on the computer, providing participants an explicit transcript that they could reflect on (Kuhn et al., 2008). The improvements in epistemic understanding of argumentation are consistent with previous empirical findings in the literature following engagement in dialogic argumentation (Jordanou, 2016b; Kuhn et al., 2013) and encountering diverse views (Barzilai & Ka'adan, 2017; Fisher et al., 2017; Kienhues et al., 2011). Yet, experimental condition participants' findings show that a combination of engagement in targeted strategies and *reflection* about their performance on those strategies proves to be a more effective way to support meta-level awareness of the norms of argumentation, thus highlighting the role of reflection in metacognitive development. Experimental condition students exhibited improvements in both their meta-strategic understanding and their epistemic understanding – the epistemic standards they employed to decide what constitute sufficient evidence to support a claim (Kuhn & Modrek, 2021).

The greater improvement in knowledge acquisition exhibited by experimental condition participants compared to control condition participants is in line with previous findings showing that engagement in reflection benefits knowledge acquisition (Bannert & Mengelkamp, 2008). The gains observed in *content knowledge* on the topic are also in line with previous work using the “Argue with me” method, where gains in knowledge acquisition were assessed by examining participants' arguments (Jordanou et al., 2019a). The present work supports those findings by employing a different method of assessing knowledge acquisition, using a multiple-choice test. Most importantly, the present work extends previous research that reported knowledge gains after participants engaged in an argumentative-based activity, by showing the unique contribution of engagement in reflective activities on argumentation in promoting knowledge gains, beyond engagement in dialogic activities. Although the alternative explanations discussed in the introduction cannot be ruled out to explain the knowledge gains, the evidence of development of participants' epistemic understanding, particularly after engagement in reflective activities, makes it fairly plausible that these gains in epistemic understanding supported knowledge gains. The development of an epistemic understanding which involves a conceptualization of knowledge as judgment, that is supported by the best available data, provides a disposition for seeking to examine and grasp a better understanding of the available data. Reflection enhanced participants' epistemic disposition, which in turn enhanced content gain.

The present findings have important educational implications, showing the effect of practice and reflection on supporting the development of argument skill on the topic examined. Previous research showed that individuals do not appreciate the role of evidence in social topics to the same degree as for physical science topics (Jordanou, 2016b). The present findings show that participants' ability to use evidence to support their critique is amenable to development when participants are offered the opportunity to engage in an argumentative-based intervention on a social topic, which is consistent with previous research (Hemberger et al., 2017; Kuhn et al., 2008). The use of the topic of immigration, in particular, that is a *controversial* social topic, shows that engagement in dialogic argumentation on controversial topics might be a promising way to develop individuals' reasoning on controversial topics, for which individuals are more prone to exhibit ‘my-side bias’ (Jordanou et al., 2020). Finally, the present findings show that engagement in practice in argument skill is a promising way to promote strategic and meta-level understanding of argumentation, but engagement in both practice and *reflection*, is an even more powerful method to support both.

## Declarations

**Informed consent** The author gained informed consent from participants.

**Disclosure of potential conflicts of interest** The author declares that she has no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Andriessen, J., & Baker, M. (2014). Arguing to learn. In R. Sawyer (Ed.), *The Cambridge Handbook of the learning sciences (cambridge handbooks in psychology)* (pp. 439–460). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139519526.027>
- Arvidsson, T. S., & Kuhn, D. (2021). Realizing the full potential of individualizing learning. *Contemporary Educational Psychology*, 65, 101960.
- Asterhan, C., & Schwarz, B. (2016). Argumentation for learning: Well-trodden paths and unexplored territories. *Educational Psychologist*, 51, 164–187.
- Bannert, M., & Mengelkamp, C. (2008). Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted. Does the verbalisation method affect learning? *Metacognition and Learning*, 3, 39–58.
- Barzilai, S., & Zohar, A. (2012). Epistemic thinking in action: Evaluating and integrating online sources. *Cognition and Instruction*, 30(1), 39–85.
- 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.
- Baytelman, A., Iordanou, K., & Constantinou, C. P. (2020). Epistemic beliefs and prior knowledge as predictors of the construction of different types of arguments on socioscientific issues. *Journal of Research in Science Teaching*, 57(8), 1199–1227. <https://doi.org/10.1002/tea.21627>
- Bereiter, C., & Scardamalia, M. (2018). Fixing Humpty Dumpty: Putting higher-order skills and knowledge together again. In L. Kerslake & R. Wegerif (Eds.), *Theory of teaching thinking: International perspective*. Routledge.
- Billig, M. (1987). *Arguing and thinking: A rhetorical approach to social psychology*. Cambridge University Press.
- Bråten, I., Britt, M. A., Strømsø, H. I., & Rouet, J.-F. (2011). The role of epistemic beliefs in the comprehension of multiple expository texts: Toward an integrated model. *Educational Psychologist*, 46, 48–70.
- Chinn, C. A., & Duncan, R. (2018). What is the value of general knowledge of scientific reasoning?. In *Scientific reasoning and argumentation: The roles of domain-specific and domain-general knowledge* (pp. 77–101). Taylor and Francis.
- Chinn, C. A., Buckland, L. A., & Samarapungavan, A. L. A. (2011). Expanding the dimensions of epistemic cognition: Arguments from philosophy and psychology. *Educational Psychologist*, 46(3), 141–167.
- Chinn, C., Barzilai, S., & Duncan, R. (2020). Disagreeing about how to know: The instructional value of explorations into knowing. *Educational Psychologist*, 55, 167–180.
- Clark, D. B., Stegmann, K., Weinberger, A., Menekse, M., & Erkens, G. (2007). Technology-enhanced learning environments to support students' argumentation. In *Argumentation in science education* (pp. 217–243). Springer.
- Cress, U. (2008). The need for considering multilevel analysis in CSCL research—An appeal for the use of more advanced statistical methods. *International Journal of Computer-Supported Collaborative Learning*, 3(1), 69–84.

- Dwyer, C. P., Hogan, M. J., & Stewart, I. (2012). An evaluation of argument mapping as a method of enhancing critical thinking performance in e-learning environments. *Metacognition and Learning*, 7(3), 219–244. <https://doi.org/10.1007/s11409-012-9092-1>
- Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist*, 46(1), 6–25.
- Felton, M. K. (2004). The development of discourse strategies in adolescent argumentation. *Cognitive Development*, 19(1), 35–52.
- Finn, B., & Metcalfe, J. (2014). Overconfidence in children's multi-trial judgments of learning. *Learning and Instruction*, 32, 1–9.
- Fisher, M., Knobe, J., Strickland, B., & Keil, F. C. (2017). The influence of social interaction on intuitions of objectivity and subjectivity. *Cognitive Science*, 41(4), 1119–1134.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, 34(10), 906.
- Flavell, J. H., & Wellman, H. M. (1975). Metamemory. Paper presented at the annual meeting of the American Psychological Association. Retrieved from <http://files.eric.ed.gov/fulltext/ED115405.pdf>.
- Grant, A. (2021). *Think again*. Viking Press
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis*. Prentice Hall.
- 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.
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures* (Vol. 22). Psychology Press.
- Iordanou, K. (2016a). 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. (2016b). Developing epistemological understanding through argumentation in scientific and social domains. *Zeitschrift Für Pädagogische Psychologie*, 30(2–3), 109–119. <https://doi.org/10.1024/1010-0652/a000172>
- Iordanou, K., & Constantinou, C. P. (2015). Supporting use of evidence in argumentation through practice in argumentation and reflection in the context of SOCRATES learning environment. *Science Education*, 99, 282–311. <https://doi.org/10.1002/sce.21152>
- 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., & Rapanta, C. (2021). “Argue with me”: A method for developing argument skills. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.631203>
- Iordanou, K., Kuhn, D., Matos, F., Shi, Y., & Hemberger, L. (2019a). Learning by arguing. *Learning and Instruction*, 63, 101–207. <https://doi.org/10.1016/j.learninstruc.2019.05.004>
- Iordanou, K., Muis, K. R., & Kendeou, P. (2019b). Epistemic perspective and online epistemic processing of evidence: Developmental and domain differences. *The Journal of Experimental Education*, 87(4), 531–551. <https://doi.org/10.1080/00220973.2018.1482857>
- Iordanou, K., Kendeou, P., & Zemblyas, M. (2020). Examining my-side bias during and after reading controversial historical accounts. *Metacognition and Learning*, 15(3), 319–342. <https://doi.org/10.1007/s11409-020-09240-w>.
- Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research* (4th ed.). Harcourt College Publishers.
- 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.
- King, P., & Kitchener, K. (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. Jossey-Bass.
- Kuhn, D. (1991). *The skills of argument*. Cambridge University Press.
- Kuhn, D. (2018). *Building our best future: Thinking critically about ourselves and our world*. Wessex Press.
- Kuhn, D. (2000). Metacognitive development. *Current Directions in Psychological Science*, 9(5), 178–181.
- Kuhn, D. (2001). How do people know? *Psychological Science*, 12(1), 1–8.
- Kuhn, D. (2020). Why is reconciling divergent views a challenge? *Current Directions in Psychological Science*, 29(1), 27–32.
- Kuhn, D., & Iordanou, K. (in press). Why do people argue past one another rather than with one another? In N. Ballantyne & D. Dunning (Eds.), *Reason, bias, and inquiry: New perspectives from the crossroads of epistemology and psychology*. Oxford University Press.

- Kuhn, D., & Modrek, A. (2021). Choose your evidence. Scientific thinking where it may most count. *Science & Education*, 11, 14. <https://doi.org/10.1007/s11191-021-00209-y>
- 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.
- 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.
- Kuhn, D., Hemberger, L., & Khait, V. (2016). Tracing the development of argumentative writing in a discourse-rich context. *Written Communication*, 33(1), 92–121.
- Lalonde, C. E., & Chandler, M. J. (2002). Children's understanding of interpretation. *New Ideas in Psychology*, 20(2–3), 163–198.
- Larrain, A., Singer, V., Strasser, K., Howe, C., López, P., Pinochet, J., Moran, C., Sánchez, Á., Silva, M., & Villavicencio, C. (2021). Argumentation skills mediate the effect of peer argumentation on content knowledge in middle-school students. *Journal of Educational Psychology*, 113(4), 736–753.
- Lin, H. S., Hong, Z. R., & Lawrenz, F. (2012). Promoting and scaffolding argumentation through reflective asynchronous discussions. *Computers & Education*, 59(2), 378–384.
- Macagno, F., Walton, D., & Reed, C. (2018). Argumentation schemes. In P. Baroni, D. M. Gabbay, M. Giacomin, & L. van der Torre (Eds.), *Handbook of formal argumentation* (pp. 517–574). College Publications.
- Mason, L., & Boscolo, P. (2000). Writing and conceptual change. What changes? *Instructional Science*, 28(3), 199–226.
- Mason, L., & Scirica, F. (2006). Prediction of students' argumentation skills about controversial topics by epistemological understanding. *Learning and Instruction*, 16(5), 492–509.
- Mason, L., Boldrin, A., & Ariasi, N. (2010). Epistemic metacognition in context: Evaluating and learning online information. *Metacognition and Learning*, 5(1), 67–90.
- Means, M. L., & Voss, J. F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. *Cognition and Instruction*, 14(2), 139–178.
- Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, 18(3), 159–163.
- Moshman, D. (2020). *Reasoning, argumentation, and deliberative democracy* (1st ed.). Routledge. <https://doi.org/10.4324/9780429316029>.
- Murphy, P. K., Greene, J. A., Firetto, C. M., Hendrick, B. D., Li, M., Montalbano, C., & Wei, L. (2018). Quality talk: Developing students' discourse to promote high-level comprehension. *American Educational Research Journal*, 55(5), 1113–1160.
- National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards*. National Governors Association Center for Best Practices & Council of Chief State School Officers.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation*, 26, 125–141.
- NGSS Lead States. (2013). *Next generation science standards: For states, by states*. The National Academies Press.
- Nussbaum, E. M., & Sinatra, G. M. (2003). Argument and conceptual engagement. *Contemporary Educational Psychology*, 28(3), 384–395.
- Nussbaum, E. M., Dove, I. J., Slife, N., Kardash, C. M., Turgut, R., & Vallett, D. (2019). Using critical questions to evaluate written and oral arguments in an undergraduate general education seminar: A quasi-experimental study. *Reading and Writing*, 32(6), 1531–1552.
- Perry, W. (1970). *Forms of intellectual and ethical development in the college years*. Holt, Rinehart, & Winston.
- Reznitskaya, A., & Gregory, M. (2013). Student thought and classroom language: Examining the mechanisms of change in dialogic teaching. *Educational Psychologist*, 48(2), 114–133. <https://doi.org/10.1080/00461520.2013.775898>
- Reznitskaya, A., 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.
- Roebbers, C. (2017). Executive function and metacognition: Towards a unifying framework of cognitive self-regulation. *Developmental Review*, 45, 31–51.
- Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, 26, 521–562.

- Ryu, S., & Sandoval, W. A. (2012). Improvements to elementary children's epistemic understanding from sustained argumentation. *Science Education*, 96(3), 488–526.
- Schneider, W., & Pressley, M. (1997). *Memory development between 2 and 20*. Erlbaum.
- Schneider, W., & Löffler, E. (2016). The development of metacognitive knowledge in children and adolescents. In J. Dunlosky & S. K. Tauber (Eds.), *Oxford library of psychology. The Oxford handbook of metamemory* (pp. 491–518). Oxford University Press.
- Shi, Y. (2019). Enhancing evidence-based argumentation in a Mainland China middle school. *Contemporary Educational Psychology*, 59, 101809.
- Shi, Y., Matos, F., & Kuhn, D. (2019). Dialog as a bridge to argumentative writing. *Journal of Writing Research*, 11(1), 107–129.
- Thomas, S. L., & Heck, R. H. (2001). Analysis of large-scale secondary data in higher education research: Potential perils associated with complex sampling designs. *Research in Higher Education*, 42(5), 517–540.
- Toulmin, S. (1958). *The uses of argument*. Cambridge University Press.
- Tsalas, N., Sodian, B., & Paulus, M. (2017). Correlates of metacognitive control in 10-year old children and adults. *Metacognition and Learning*, 12(3), 297–314.
- Van Der Veen, C., & Van Oers, B. (2017). Advances in research on classroom dialogue: Learning outcomes and assessments. *Learning and Instruction*, 48, 1–4.
- Veenman, M. V. J. (2017). Learning to self-monitor and to self-regulate. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction*. 2<sup>nd</sup> (revised ed., pp. 233–257). Routledge.
- Veenman, M. V., Van Hout-Wolters, B. H., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, 1(1), 3–14.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Walton, D. N. (1989). Dialogue theory for critical thinking. *Argumentation*, 3, 169–184.
- Walton, D. (2000). The place of dialogue theory in logic, computer science and communication studies. *Synthese*, 123(3), 327–346.
- Wecker, C., & Fischer, F. (2014). Where is the evidence? A meta-analysis on the role of argumentation for the acquisition of domain-specific knowledge in computer-supported collaborative learning. *Computers & Education*, 75, 218–228.
- Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, 46, 71–95.
- Wiley, J., Griffin, T. D., Steffens, B., & Britt, M. A. (2020). Epistemic beliefs about the value of integrating information across multiple documents in history. *Learning and Instruction*, 65, 101266.
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press.
- Zohar, A., & Ben David, A. B. (2009). Explicit teaching of meta-strategic knowledge in authentic classroom situations. *Metacognition and Learning*, 3(1), 59–82.