Science & Education (2023) 32:123–168
https://doi.org/10.1007/s11191-021-00288-x

ARTICLE

Long-Lasting Conceptual Change in Science Education

The Role of U-shaped Pattern of Argumentative Dialogue in Collaborative Argumentation

Xiaoshan Li[1] · Yanyan Li[1] · Wenjing Wang[2]

Accepted: 23 September 2021 / Published online: 16 November 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract

Meaningful learning for conceptual change in science education should aim to help students change their existing misconceptions to develop an accurate understanding of scientific concepts. Although collaborative argumentation is assumed to support such processes, its value for conceptual change is unclear. Moreover, the roles of argumentative dialogue should be considered in studies on collaborative argumentation. In the present study, using a controlled experiment, we examined the value of collaborative argumentation for conceptual change in science education while fully considering the roles of argumentative dialogue. Twenty-three postgraduate students were each allocated to one of two conditions (individual argumentation [control group] and collaborative argumentation [experimental group]) and participated in two argumentation activities. The results revealed that collaborative argumentation had a delayed but long-lasting effect on conceptual change in science education (i.e., conceptual change induced by collaborative argumentation did not immediately indicate a significant improvement at the moments of argumentation but showed a significant improvement during the delay period). Collaborative argumentation provided opportunities for change in cognitive, ontological, intentional, and other aspects of learning. Dialogue protocol analysis revealed that long-lasting conceptual change was associated with a U-shaped pattern of argumentative dialogue (i.e., two high and one low: both deliberative argumentation and co-consensual construction frequently occurred, while disputative argumentation rarely occurred) in collaborative argumentation. A third argumentation activity was then conducted to confirm this unexpected finding. The results confirmed an association between long-lasting conceptual change and a U-shaped pattern of argumentative dialogue in collaborative argumentation. The current study sheds light on the value of collaborative argumentation for long-lasting conceptual change, deepening our understanding of whether conceptual gains from argumentation activities were contingent on a particular type of verbal dialogue powered by collaborative argumentation. Implications for science education were discussed.

✉ Yanyan Li
liyy@bnu.edu.cn

Extended author information available on the last page of the article

1 Introduction

Promoting scientific literacy among students is generally considered to be the central purpose of science education (NGSS, 2013; NRC, 2012). Students with more highly developed scientific literacy demonstrate a greater ability to accurately understand and use scientific concepts to generate explanations or make predictions about natural phenomena (Bybee, 2008). Science educators and practitioners have reported that it can be difficult to help students change their existing misconceptions to develop an accurate understanding of scientific concepts (i.e., conceptual change) because many misconceptions are robust and resistant to educational interventions and teaching approaches (Anderson & Smith, 1987; Asterhan & Dotan, 2018; Chi, 2005).

In the last 20 years, argumentation, as one of the science education practices advocated by the Next Generation Science Standards (NGSS, 2013; NRC 2012), has received increasing interest because of its effects on science content learning (Asterhan & Resnick, 2020; Driver et al., 2000; Erduran, 2007).

Within approaches in which students construct arguments, there are two main types of argumentation: individual argumentation and collaborative argumentation (Kilinc et al., 2017; Walton, 2009). Individual argumentation concerns the way in which an individual constructs an argument (Goldman, 1999), focusing on constructing individual knowledge by presenting arguments in support of the thesis (Reed & Long, 1998). Collaborative argumentation refers to dialogical argumentation that takes place in groups of students when they are asked to work together on a common task of constructing an argument (Asterhan & Schwarz, 2007; Evagorou & Osborne, 2013). Collaborative argumentation emphasizes social interaction, intellectual openness, deep-level joint thinking (Isohätälä et al., 2018), and social co-construction of scientific knowledge (Evagorou & Osborne, 2013; Walton, 2009). As suggested in Mclure et al. (2020)'s multidimensional framework of conceptual change and previous studies of argumentation, collaborative argumentation might be a promising approach to promote conceptual change through making changes in cognitive, ontological, and intentional dimensions (Asterhan & Schwarz, 2016; McLure et al., 2020). Especially for making changes in the intentional dimension, collaborative argumentation could provide a highly open environment with students in which they could not only examine their beliefs and motivations through comparison with others but also be inspired and encouraged by others' strong beliefs and motivations (Inagaki & Hatano, 2003; Vosniadou, 2003).

However, very few empirical studies have explicitly examined the value of collaborative argumentation for conceptual change in science education (Asterhan & Resnick, 2020; Asterhan & Schwarz, 2016). Moreover, recent studies have highlighted a paradox: collaborative argumentation may not necessarily lead to productive outcomes if the factor of argumentative dialogue in collaborative argumentation is not taken into consideration (Asterhan & Babichenko, 2015; Asterhan & Schwarz, 2016; Liu et al., 2019; Yang et al., 2015). In other words, merely focusing on argumentation activities per se was not enough to gain a deep understanding of the value of collaborative argumentation for conceptual change. We need to go further and scrutinize a crucial factor—argumentative dialogue that occurred in the processes of collaborative argumentation. Therefore, by conducting comparisons with individual argumentation, this study aimed to explore the value of collaborative argumentation for conceptual change in science education when comprehensively considering the roles of argumentative dialogue.

2 Literature Review

2.1 Theoretical Framework

2.1.1 Conceptual Change in Science Education

Chi (2008) argued that "conceptual change is not *adding* new knowledge or *gap filling* incomplete knowledge; rather, conceptual change is changing misconceptions in existing knowledge structures to an accurate understanding of scientific concepts". In other words, conceptual change occurs when students come to have an improved capability to construct and identify a scientifically accurate and full scientific explanation, involving a substantive reorganization or revisions of their existing knowledge structures that were embedded with misconceptions (Asterhan & Resnick, 2020; Asterhan & Schwarz, 2016; Duit & Treagust, 2003, 2012). Many previous studies have reported that it is difficult to achieve conceptual change in science education because some misconceptions tend to be stable, robust, and resistant to standard tell-and-practice teaching approaches and even innovative instructional interventions (Anderson & Smith, 1987; Asterhan & Dotan, 2018; Asterhan & Schwarz, 2007; Champagne et al., 1985; Chi, 2008; Dreyfus et al., 1990; Hake, 1998; Jensen & Finley, 1996; Jiménez-Aleixandre, 1992; Vosniadou & Mason, 2012). Moreover, determining how to maintain conceptual change over the long term is an emergent problem in science education (Asterhan & Resnick, 2020; Kaya, 2013; McLure et al., 2020).

Misconceptions have been considered as "theory-like naïve assumptions held by medieval scientists" (Chi, 2005) and as "a set of loosely connected and reinforcing ideas" (diSessa, 1988). Previous studies have proposed several explanations for the robustness of such misconceptions (Chi, 2005, 2008). First, with significant breakthroughs in research technologies, such as cryo-electron microscopy (cryo-EM) and polymerase chain reaction (PCR), scientific concepts are often complex and counter-intuitive, contradicting students' intuitive understanding because the world of modern science is removed from the everyday world we experience. Students may naively conceive of science-related concepts as one type of concept when in fact they are another (Ferrari & Chi, 1998), such as initially conceiving of concepts in physics, such as heat, electrical current, and light, as a type of substance (Chi, 1992, 1997; Chi et al., 1994) or conceiving of diffusion as a type of event. In fact, all of these science-related concepts are equilibration processes (Ferrari & Chi, 1998). In other words, there may be an incongruence between students' intuitive understanding and scientists' explanatory schemata (Shtulman, 2006). Second, the relationships among scientific concepts are multifaced and systematically connected, constituting a "group" or "cluster" (Johnstone, 2000; Taber, 2013; Talanquer, 2011). The relationships among these concepts exhibit a triangular structure that can be described as a triangle with three apices labeled "macro", "symbolic", and "sub-micro" representations (Johnstone, 2000; Taber, 2013; Talanquer, 2011). The first term refers to macro representations of things that can be seen, touched, and smelled, such as substances, chemical reactions, animals, and plants. The second term refers to sub-micro representations of things that can only be tested indirectly, for which meaning cannot be derived from direct observation. Molecules, ions, genes, and electrical interactions are classified as belonging to this level. The third term refers to symbolic representations, including formulae, equations, and mathematical manipulation and graphs, which act as a bridge between the aforementioned two levels by simultaneously representing both macro and sub-micro representations, as well as aiding students in shifting between these levels (Heng et al., 2015;Johnstone, 2000 ; Taber, 2013).

For example, in the case of sodium chloride dissolving, the macro representation refers to the phenomenon that salt dissolves in water, the sub-micro representation refers to the phenomenon that sodium chloride existing in form of a regular lattice is attracted to water molecules and is towed off into the solution, and the symbolic representation refers to the formulae "$Na^+ Cl^- (s) + H_2O \rightarrow Na^+ (aq) + Cl^- (aq)$" (Johnstone, 1991). In current science education, educators and teachers commonly present scientific concepts to students with the three levels of representations, rather than supporting students to develop multileveled ways of thinking (Johnstone, 2000; Taber, 2013). Johnstone (2000) argues that this situation has provided the origins of robust misconceptions. Thus, many science educators and practitioners have called for a powerful means to facilitate conceptual change in science education (Chi, 2008; McLure et al., 2020).

2.1.2 Collaborative Argumentation for Conceptual Change

Argumentation is one of the science practices included in the Next Generation Science Standards (NGSS, 2013; NRC 2012), which have received increased interest in the last 20 years because of its effects on science content learning (Asterhan & Schwarz, 2016; Asterhan & Resnick, 2020; Driver et al., 2000; Erduran, 2007; Jimenez-Aleixandre and Erduran, 2007; Katchevich et al., 2013; Liu et al., 2019; Yang et al., 2015). Scholars examining argumentation have categorized argumentation into multiple principles (Liu et al., 2019). Based on approaches regarding the ways in which students construct an argument, argumentation can be categorized into two types: individual argumentation and collaborative argumentation (Asterhan & Schwarz, 2007; Kilinc et al., 2017; Liu et al., 2019; Ryu & Sandoval, 2008; Sampson & Clark, 2009; Walton, 2009). Individual argumentation relates to the way in which an individual constructs an argument (Goldman, 1999), and is also referred to as monological argumentation, involving "implicit dialogue" (van Eemeren and Grootendorst 1984, p.12). Individual argumentation is not merely an individual soliloquy or "chain of individual reasoning" (Reed & Long, 1998, p.2), but rather focuses on the intuitive "case building" of presenting arguments in support of a thesis (Reed & Long, 1998, p.2). For example, Charles Darwin describes his book *On the Origin of Species* as consisting of one long argument (Jiménex-Aleixandre and Erduran, 2007, p. 3). He presents his claim and scientific discovery through converging lines of reasoning, theoretical idea, and empirical evidence. Piaget's theory of constructivism theoretically explains how individual students engage in individual argumentation and construct individual knowledge by connecting present evidence to individual experience in one's mind (Piaget, 1954). The second type of argumentation, collaborative argumentation, refers to dialogical argumentation that takes place among groups of students when they are asked to work together on a common task of constructing and presenting an argument (Asterhan & Schwarz, 2007; Evagorou & Osborne, 2013). Collaborative argumentation emphasizes participants' social interaction, intellectual openness, deep-level joint thinking (i.e., group members' mutual engagement in joint discussions to reach joint understandings and well-reasoned decisions) (Isohätälä et al., 2018), and social co-construction of scientific knowledge (Evagorou & Osborne, 2013; Walton, 2009).

Recent studies of conceptual change in science education indicated a trend toward a multidimensional perspective (Chi, 2008; Duit & Treagust, 2003, 2012; McLure et al., 2020; Posner et al., 1982; Sinatra, 2003; Sinatra & Pintrich, 2003; Vosniadou, 2003). The multidimensional perspective of conceptual change provides an explanation for why certain teaching approaches can or cannot promote conceptual change in science education

(McLure et al., 2020). Through synthesizing and integrating previous theories or frameworks on conceptual change, McLure et al. (2020) established a multidimensional conceptual change framework—conceptual change should focus on not only change in cognitive aspects of learning but also consider change in ontological and intentional aspects of learning. Based on Mclure et al. (2020)'s multidimensional conceptual change framework and previous studies of argumentation, it suggested that collaborative argumentation might be a promising approach for promoting conceptual change through supporting change in cognitive, ontological, and intentional aspects of learning (Asterhan & Resnick, 2020; Asterhan & Schwarz, 2016; Heng et al., 2015; Liu et al., 2019; McLure et al., 2020; Sampson & Clark, 2009, 2011; Yang et al., 2015). In collaborative argumentation, students engaged in a host of activities associated with developing a deep understanding of domain-specific content (Asterhan & Schwarz, 2016).

First, in Mclure et al. (2020)'s multidimensional conceptual change framework, cognitive aspects of learning refer to change in conceptual contents. Many studies have reported that collaborative argumentation provides opportunities for promoting change in cognitive aspects of learning (Asterhan & Resnick, 2020; Asterhan & Schwarz, 2016; Heng et al., 2015; Sampson & Clark, 2009, 2011). Through being exposed to evaluating and comparing alternative ideas articulated by group members, students actively begin to perceive, self-reflect, and diagnose the misconceptions in their original ideas and then reorganize and remediate these misconceptions to generate more sound scientific explanations for discrepant phenomena (Asterhan & Resnick, 2020; Asterhan & Schwarz, 2016; Heng et al., 2015;Sampson & Clark, 2009, 2011). This process involves an iterative pathway, in which students repeatedly scrutinize and improve their preceding intuitive naïve notion, developing a more scientifically valid account.

Second, several scholars have reported that collaborative argumentation can also provide opportunities for facilitating ontological dimension of conceptual change (Asterhan & Resnick, 2020; Asterhan & Schwarz, 2016). Change in ontological aspects of learning includes change in conceptual models within a specific disciplinary context or change in understanding of the nature of science (Duit & Treagust, 2003; McLure et al., 2020). The term "ontological" is used to explain changes to the way students conceptualize science entities (Chi, 2008; Chi et al., 1994; Duit & Treagust, 2003; Thagard, 1992; Vosniadou, 1994). When given the same science concept, students' conceptualization process is incommensurable with scientists' conceptualization process (Chi et al., 1994; Chi 2008; Duit and Treagust 2003). For example, students described heat as a flowing fluid and described the gene as an inherited object, while scientists viewed heat as kinetic energy in transit and viewed the gene as a biochemical process (Duit and Treagust 2010). Change from students' material conceptions to scientists' process view was a good example of conceptual change at the ontological level (Duit & Treagust, 2003). When engaging in argumentation-based activities, students can revise and restructure their own conceptual models by being exposed to and comparing scientists' or others' ways of thinking. Ontological conceptual change can be a revolutionary change in students' conceptual models, rather than an incremental change in students' epistemological restructuring (Vosniadou, 2003). Such a change can result in the "tree swapping" (Chi et al., 1994) or "tree switching" effect (Thagard, 1992).

Finally, some researchers have emphasized an emergent focus on the benefits of collaborative argumentation for change in intentional aspects of learning (Asterhan & Resnick, 2020; Asterhan & Schwarz, 2016). And recent studies have reported that changes in cognitive and ontological dimensions have limited effects if science educators and practitioners disregard intentional dimension (McLure et al., 2020). Based on McLure et al. (2020)'s multidimensional conceptual change framework, change in intentional aspects of learning consisted of change at levels of personal attitudes and beliefs related to science learning. Students who engage in intentional learning tend to be aware of their goals and beliefs, internally oriented to goals and deliberate actions, motivated to focus on tasks, and are willing to restructure their understanding under their own conscious control (Luque, 2003). Even when encountering the most difficult and counter-intuitive concepts of modern science, such students can persist in dealing with these problems through intentionally taking multiple cognitive and metacognitive strategies rather than being passively controlled by the level of difficulty of the task (Sinatra, 2003). Collaborative argumentation does not merely provide opportunities for social interaction and collaboration (Sampson & Clark, 2009, 2011) but could also create a highly open environment that elicits a large amount of discussion and debate (Asterhan & Schwarz, 2016;Inagaki & Hatano, 2003 ; Vosniadou, 2003). This highly open environment may allow students to not only examine their beliefs and motivations through comparison with others, but also to be inspired and encouraged by others' strong beliefs and motivations, becoming more engaged and persistent in science learning when they are met with challenges and difficulties (Inagaki & Hatano, 2003; Vosniadou, 2003). Such changes in intentional aspects are expected to result in resilient scientific concept learning and long-lasting conceptual change (McLure et al., 2020; Sinatra, 2003; Sinatra & Taasoobshirazi, 2011).

Scholars of argumentation have discerned the promising benefits of collaborative argumentation for conceptual change, but very few empirical studies have explicitly stated the value of collaborative argumentation in terms of conceptual change (Asterhan & Resnick, 2020; Asterhan & Schwarz, 2016). Therefore, some scholars have called for more empirical studies to examine whether and why collaborative argumentation promotes conceptual change in science education (Asterhan & Resnick, 2020; Asterhan & Schwarz, 2016).

2.1.3 Argumentative Dialogue in Collaborative Argumentation

Recent studies have highlighted a paradox: collaborative argumentation does not necessarily lead to productive outcomes if the factor of argumentative dialogue in collaborative argumentation is not taken into consideration (Asterhan & Babichenko, 2015; Asterhan & Resnick, 2020; Asterhan & Schwarz, 2016; Liu et al., 2019; Yang et al., 2015). For example, Liu et al. (2019) conducted a quasi-experiment to investigate the influence of online collaborative argumentation on students' scientific concept learning among 7th grade students, using scores on pre- and post-tests of scientific concept assessment in an activity. Students in the experimental group engaged in a collaborative version of online argumentation, while students in the control group engaged in an individual version of online argumentation. The results revealed that the type of argumentation approach did not influence students' outcomes regarding scientific concept learning. However, further analysis of the argumentation process revealed that some students who engaged in the collaborative version of online argumentation misused critiques or rebuttals for disputative goals (e.g., to win the argument) or used personal attacks in the argument, rather than pursuing the goal of knowledge co-construction (Mercer, 2000). These findings indicate that not all dialogue in collaborative argumentation promotes learning, and some may act as barriers to productive outcomes (Heng et al., 2015; Osborne et al., 2004).

Another previous study articulated a need to further investigate verbal dialogue in dyadic argumentation, based on results revealing no differences between students in the conditions of dyadic argumentation and students in the conditions of individual problem solving in terms of scores of conceptual understanding of biological evolution (Asterhan & Resnick, 2020). According to the sociocultural theory proposed by Vygotsky (1978), individuals' cognition is shaped through social interactions, and dialogue plays a special role in this process. Accordingly, when students engage in dialogue activities that require them to articulate incomplete ideas, to examine their misunderstandings, or to challenge or be challenged by peers, they may be more able to process the domain-specific content of a given topic, resulting in more effective conceptual change (Asterhan & Schwarz, 2016). Thus, some scholars of argumentation have argued that more attention should be paid to the special roles of argumentative dialogue in collaborative argumentation (Asterhan & Resnick, 2020; Asterhan & Schwarz, 2016; Liu et al., 2019).

Based on previous psychoeducational studies, Asterhan and Schwarz (2016) reviewed three types of argumentative dialogue in learning contexts: deliberative argumentation, disputative argumentation, and consensual co-construction. The first type, deliberative argumentation (Felton et al., 2009), is commonly believed to be an idealized form of argumentative dialogue for learning, although some inevitable differences among scholars' descriptions of deliberative argumentation exist (Asterhan and Schwarz, 2009; Berland & Hammer, 2012; Felton et al., 2009; Mercer, 1996; Nussbaum, 2008). Deliberative argumentation is a type of dialogue that maintains a balance between critical reasoning and collaborative knowledge construction and is characterized by the following characteristics (Asterhan, 2013): (1) the dialogue between discussants showed a willingness to be a good listener, to be a critical thinker, and to be an idea hunter when alternative perspectives that had not yet been considered were proposed by peers; (2) the dialogue between discussants showed a willingness to make concessions in response to sound or cogent arguments; (3) this dialogue was not *position*-driven but *issue*-driven. Even when there was disagreement between discussants, there were no manifest expressions of discomfort or interpersonal tension in the interaction. Further elaborations were requested, and the discussants attempted to understand each other.

The second type of argumentative dialogue refers to disputative dialogue, in which discussants defended a viewpoint and undermined alternative viewpoints to convince opponents to switch sides (Asterhan & Schwarz, 2016). Besides being rich in critical reasoning, disputative argumentation had the following characteristics: (1) the dialogue between discussants showed a focus not on the collaborative co-construction of knowledge, but on the interpersonal competitive dimension of social interaction, wherein discussants incorrectly viewed scientific argumentation as a "win-lose" activity (Asterhan & Babichenko, 2015). (2) Perceptions of interpersonal competition might not only raise concerns about interpersonal relationships or senses of group belonging (Asterhan & Schwarz, 2016), but may also reduce cognitive flexibility and a person's openness to alternative viewpoints (Carnevale & Probst, 1998). (3) Disputative argumentation was likely to deteriorate into "appeal to force" or "personal attack in argument" (Woods, 2004), causing discussants to concede upfront without further consideration and engagement (Asterhan & Schwarz, 2016; Smith et al., 1981; Weinberger & Fischer, 2006). The third type of argumentative dialogue is consensual co-construction, in which discussants transacted on each other's contributions by agreeing with, elaborating on and expanding on ideas (Asterhan & Schwarz, 2016). This dialogue between discussants showed mere development and consolidation of one-sided arguments, rather than challenging or juxtaposing different alternatives.

2.2 Related Studies

Some previous studies attempted to explore the association between argumentative dialogue and conceptual understanding of scientific concepts. For example, Asterhan and Babichenko (2015) conducted a strictly controlled study to examine whether particular types of argumentative dialogue are associated with better conceptual understanding of scientific concepts (i.e., diffusion). Each student was instructed to pair up and interact with a virtual or real confederate in a computer-mediated online context. The confederate's verbal behavior was scripted to evoke argumentative dialogue, while controlling exposure to the content of scientific concepts and the type of dialogue (e.g., requests for clarifications, challenges), but differing in the types of argumentative dialogue (i.e., disputative or deliberative). The results revealed that individuals who participated in the deliberative dialogue condition outperformed those in the disputative dialogue condition on conceptual understanding test scores. Examination of the dialogue revealed that individuals in the deliberative dialogue condition more openly shared their incomplete understanding with confederates. Although these findings confirmed the expectation that deliberative argumentation was associated with better conceptual understanding, the authors proposed that instead of *assuming* that students participate in a particular type of argumentative dialogue (e.g., because we told them so, or because we expected them to), it is imperative to carefully describe and examine the *actual* argumentative dialogue that ensued, such as in the context of collaborative argumentation (Asterhan & Resnick, 2020; Asterhan & Schwarz, 2016). Moreover, it was not sufficient to focus on a particular type of argumentative dialogue, because distinct argumentative dialogue was likely to offer different opportunities for learning, which would be expected to affect learning outcomes differently (Asterhan & Resnick, 2020; Asterhan & Schwarz, 2016). Therefore, some scholars have proposed the need for studies of collaborative argumentation that comprehensively consider the roles of argumentative dialogue (Asterhan & Schwarz, 2016).

3 The Purpose of the Research

The previous studies reviewed above indicate that collaborative argumentation could provide a promising approach for promoting conceptual change in science education. Moreover, the role of argumentative dialogue should be considered in studies on collaborative argumentation. Therefore, the aim of the current study was to examine the value of collaborative argumentation for long-lasting conceptual change. To achieve this goal, the following research questions were investigated:

1) Do students who engage in collaborative argumentation demonstrate better conceptual change compared with students who engage in individual argumentation?
2) What opportunities could collaborative argumentation offer for change in cognitive, ontological and intentional aspects of learning?
3) Is long-lasting conceptual change associated with a particular pattern of argumentative dialogue in collaborative argumentation?

4 Methodology

This was a mixed-method study, including qualitative research and quantitative research. We added a brief table to make clear which research paradigms, which research methods, which data sources and which data analysis were used for addressing each research question in this study (see Table 1).

4.1 Participants

Twenty-three second-year postgraduate students (mean age = 24.83 years, SD = 1.90; one male, 22 female) at a large comprehensive university in central China participated in this study. Twenty-two students were Chinese and one student came from the Republic of Kazakhstan. The participants all had the same major (science education) and were in the same class. In accordance with the Master's Degree of Science Education issued by the Ministry of Education of China and the education talents training plan issued by the university, science education postgraduate students were required to be well-prepared with science domain-specific knowledge and pedagogical knowledge of science education. A series of curricula on science education were taught in the spring semester of the 2019–2020 academic year. One compulsory course was *Educational Technology and the Development of Science Education*, which aimed to briefly introduce and let students experience how a variety of educational technologies have been used to facilitate and support science teaching and learning in the past several decades. Therefore, argumentation as an important means of educational technology was introduced in this course.

This course began on 1 March 2020 and lasted for 14 weeks. The course involved a total of 32 study hours. Because of significant disruption caused by the COVID-19 pandemic, the *Educational Technology and the Development of Science Education* class was shifted from a face-to-face context to an online context. The teacher who instructed this course had proficient skills in integrating information technology with the requirements of teaching. By using information technology platforms, including Zoom Rooms, Star of Questionnaires (an online platform to manage, conduct and collect questionnaires according to clients' requirements), Collaborative Editing Platform and Tencent QQ, the whole process of teaching and learning, which included teacher instruction, completing conceptual understanding tests and questionnaires, participating in argumentation activities and carrying out interviews, was implemented in the online context.

Prior to the study, each student was instructed to participate in a simple interview for 10 min, inquiring about whether they had any previous experience of argumentation in the context of science education or other contexts, which topics in high school biology they were interested in. At the end of the simple interview, all of the students were informed about the main objective and procedures of this study in detail and provided informed consent to record their speaking and images. Participants were told that they could refuse to be recorded and could leave the study at any time. They were also told that their information, recordings, and scores on conceptual understanding would only be used for the purposes of the research and would in no way influence their marks in the curriculum at the end of the school year.

Table 1 Research paradigms, research methods, data sources and data analysis used for addressing each research question in this study

Research questions	Research paradigm	Research methods	Data sources	Data analysis
Research question 1	Quantitative research	Experimental method	Pre-test, Immediate post-test, and delayed post-test of conceptual understanding	Scoring of tests and statistics analysis
Research question 2	Qualitative research	Phenomenological method	Stimulated recall interviews	Coding of interviews and counting of categories' frequencies
Research question 3	Quantitative research	Correlational method	Dialogues within collaborative groups	Coding of dialogues and statistics analysis

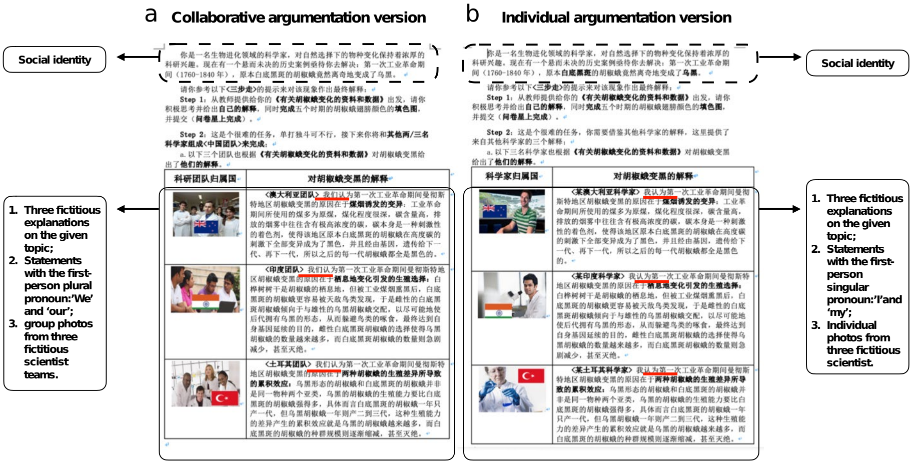
4.2 Topics and Questions for Three Argumentation Activities

Three topics were selected from a list of topics that was involved in the National Biology Curriculum Standard for High School in China, since most of students stated they were interested in them in the simple interview. Task designs for argumentation activities surrounded the three topics, aiming to create a problematized but scaffolded environment where students could engage in argumentation.

The first topic for argumentation was called *A Phantom of the Industrial Revolution.* This topic referred to a bizarre phenomenon during the Industrial Revolution in England, in which peppered moths that previously had light-colored wing patterns became blackened. The evolution of the peppered moth was an effective teaching example of Darwin's theory of *Natural Selection*. Because none of the participants had a background in life sciences in higher education, background materials on this topic had been prepared for students to gain an overview of this event, including various images of peppered moths and their habitat, a description of the resting behavior of peppered moths, and population data of peppered moths at five periods of the Industrial Revolution (see Appendix 5). Two versions of worksheets were prepared for guiding argumentation activities. The collaborative argumentation version was prepared for students in the collaborative argumentation condition, and the individual argumentation version was prepared for students in the individual argumentation condition. Each participant received the same social identity (e.g., a biological scientist), instructions of "Three Steps Ready Go", and three fictitious explanations from three fictitious teams (or individuals) of scientists. The only difference between the two versions of the worksheets was that the first-person plural pronoun (i.e., "we") and possessive determiner (i.e., "our") were used in the collaborative argumentation version to create a sense of collaboration in the collaborative argumentation condition. In contrast, the first-person singular pronoun (i.e., "I") and possessive determiner (i.e., "my") were used in the individual argumentation version. In addition, group photos from three fictitious teams of scientists and individual photos from three fictitious scientists were presented in the collaborative and individual argumentation versions of worksheets, respectively.

Examples of the two versions of the worksheet are presented and compared in Fig. 1, and the complete English versions of the worksheets are also presented in Appendix 1 and Appendix 2. The following starting question for students to begin argumentation in both the Star of Questionnaires and Collaborative Editing Platform was presented: "Please attempt to give explanations for why the light-colored peppered moths became blackened during the Industrial Revolution in England". Moreover, students were instructed to complete a coloring task that graphically depicted the gradual change in wing coloring in the five periods of the Industrial Revolution (see Fig. 2). And the complete English version of the coloring task is also presented in Appendix 4. The coloring task was adapted from Shultman's (2006) study and was designed to provide an opportunity for students to visualize their explanations of peppered moth evolution. Whether students accurately understood scientific concepts of natural selection was examined by analyzing their graphical explanations. Students drew five simple circles to represent peppered moths and used a pencil to shade the circles to represent the magnitude of change in peppered moth coloring. When students had completed the coloring task, they took photos of their work and uploaded the photos to the Star of Questionnaires platform.

The second topic for argumentation was *Jiankui He in the Eye of the Storm.* Jiankui He was a Chinese researcher who produced the first genetically edited human babies in



Collaborative argumentation version

Social identity

1. Three fictitious explanations on the given topic;
2. Statements with the first-person plural pronoun:'We' and 'our';
3. group photos from three fictitious scientist teams.

你是一名生物进化领域的科学家，对自然选择下的物种变化保持着浓厚的科研兴趣。现在有一个悬而未决的历史案例亟待你去解决：第一次工业革命期间（1760-1840 年），原本白底黑斑的胡椒蛾竟然离奇地变成了乌黑。

请你参考以下**<三步走>**的提示来对该现象作出最终解释：

Step 1：从教师提供给你的**《有关胡椒蛾变化的资料和数据》**出发，请你积极思考并给出**自己的解释**，同时**完成**五个时期的胡椒蛾翅膀颜色的**填色图**，并提交（**问卷星上完成**）。

Step 2：这是个很难的任务，单打独斗可不行，接下来你将和**其他两/三名科学家组成<中国团队>来完成**：

a. 以下三个团队也根据**《有关胡椒蛾变化的资料和数据》**对胡椒蛾变黑给出了**他们的解释**。

科研团队归属国	对胡椒蛾变黑的解释
	<澳大利亚团队> 我们认为第一次工业革命期间曼彻斯特地区胡椒蛾变黑的原因在于**煤烟诱发的变异**：工业革命期间所使用的煤多为原煤，煤化程度很深，碳含量高，排放的烟雾中往往含有极高浓度的碳，碳本身是一种刺激性的着色剂，使得该地区原本白底黑斑的胡椒蛾在高度碳的刺激下全部变异成为了黑色，并且经由基因，遗传给下一代、再下一代，所以之后的每一代胡椒蛾都全是黑色的。
	<印度团队> 我们认为第一次工业革命期间曼彻斯特地区胡椒蛾变黑的原因在于**栖息地变化引发的生殖选择**：白桦树树干是胡椒蛾的栖息地，但被工业煤烟熏黑后，白底黑斑的胡椒蛾更容易被天敌鸟类发现，于是雌性的白底黑斑胡椒蛾倾向于与雄性的乌黑胡椒蛾交配，以尽可能地使后代拥有乌黑的形态，从而躲避鸟类的啄食，最终达到自身基因延续的目的，雌性白底黑斑胡椒蛾的选择使得乌黑胡椒蛾的数量越来越多，而白底黑斑胡椒蛾的数量则急剧减少，甚至灭绝。
	<土耳其团队> 我们认为第一次工业革命期间曼彻斯特地区胡椒蛾变黑的原因在于**两种胡椒蛾的生殖差异所导致的累积效应**：乌黑形态的胡椒蛾和白底黑斑的胡椒蛾并非是同一物种两个亚类，乌黑的胡椒蛾的生殖能力要比白底黑斑的胡椒蛾强得多，具体而言白底黑斑的胡椒蛾一年只产一代，但乌黑胡椒蛾一年则产二到三代，这种生殖能力的差异产生的累积效应就是乌黑的胡椒蛾越来越多，而白底黑斑的胡椒蛾的种群规模则逐渐缩减，甚至灭绝。

Individual argumentation version

Social identity

1. Three fictitious explanations on the given topic;
2. Statements with the first-person singular pronoun:'I'and 'my';
3. Individual photos from three fictitious scientist.

你是一名生物进化领域的科学家，对自然选择下的物种变化保持着浓厚的科研兴趣。现在有一个悬而未决的历史案例亟待你去解决：第一次工业革命期间（1760-1840 年），原本**白底黑斑**的胡椒蛾竟然离奇地变成了**乌黑**。

请你参考以下**<三步走>**的提示来对该现象作出最终解释：

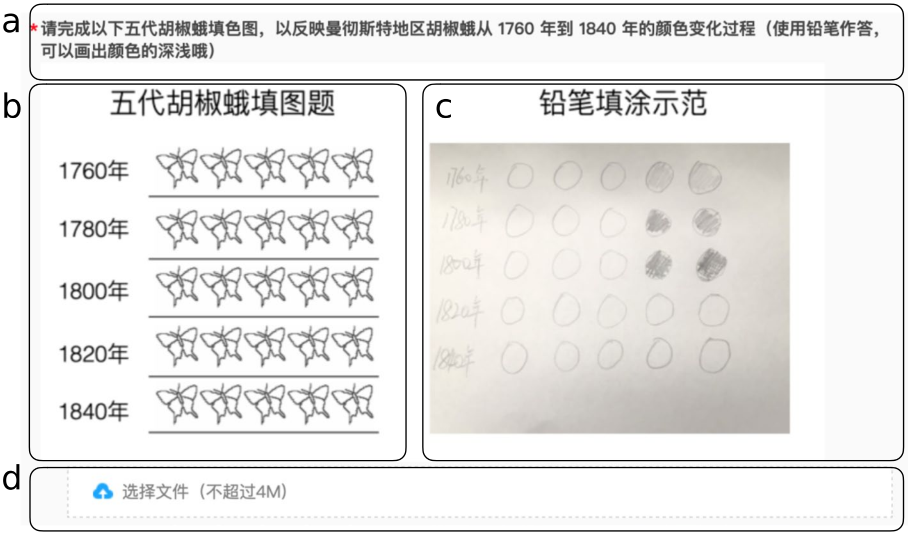
Step 1：从教师提供给你的**《有关胡椒蛾变化的资料和数据》**出发，请你积极思考并给出**自己的解释**，同时**完成**五个时期的胡椒蛾翅膀颜色的**填色图**，并提交（**问卷星上完成**）。

Step 2：这是个很难的任务，你需要借鉴其他科学家的解释，这里提供了来自其他科学家的三个解释：

a. 以下三名科学家也根据**《有关胡椒蛾变化的资料和数据》**对胡椒蛾变黑给出了**他们的解释**。

科学家归属国	对胡椒蛾变黑的解释
	<某澳大利亚科学家> 我认为第一次工业革命期间曼彻斯特地区胡椒蛾变黑的原因在于**煤烟诱发的变异**：工业革命期间所使用的煤多为原煤，煤化程度很深，碳含量高，排放的烟雾中往往含有极高浓度的碳，碳本身是一种刺激性的着色剂，使得该地区原本白底黑斑的胡椒蛾在高度碳的刺激下全部变异成为了黑色，并且经由基因，遗传给下一代、再下一代，所以之后的每一代胡椒蛾都全是黑色的。
	<某印度科学家> 我认为第一次工业革命期间曼彻斯特地区胡椒蛾变黑的原因在于**栖息地变化引发的生殖选择**：白桦树树干是胡椒蛾的栖息地，但被工业煤烟熏黑后，白底黑斑的胡椒蛾更容易被天敌鸟类发现，于是雌性的白底黑斑胡椒蛾倾向于与雄性的乌黑胡椒蛾交配，以尽可能地使后代拥有乌黑的形态，从而躲避鸟类的啄食，最终达到自身基因延续的目的，雌性白底黑斑胡椒蛾的选择使得乌黑胡椒蛾的数量越来越多，而白底黑斑胡椒蛾的数量则急剧减少，甚至灭绝。
	<某土耳其科学家> 我认为第一次工业革命期间曼彻斯特地区胡椒蛾变黑的原因在于**两种胡椒蛾的生殖差异所导致的累积效应**：乌黑形态的胡椒蛾和白底黑斑的胡椒蛾并非是同一物种两个亚类，乌黑的胡椒蛾的生殖能力要比白底黑斑的胡椒蛾强得多，具体而言白底黑斑的胡椒蛾一年只产一代，但乌黑胡椒蛾一年则产二到三代，这种生殖能力的差异产生的累积效应就是乌黑的胡椒蛾越来越多，而白底黑斑的胡椒蛾的种群规模则逐渐缩减，甚至灭绝。

Fig. 1 Two versions of worksheets. (**a**) Collaborative argumentation version of worksheet; (**b**) individual argumentation version of worksheet. Full lines indicate the difference between the two versions of worksheets, and dotted lines indicate the similarity



*请完成以下五代胡椒蛾填色图，以反映曼彻斯特地区胡椒蛾从 1760 年到 1840 年的颜色变化过程（使用铅笔作答，可以画出颜色的深浅哦）

五代胡椒蛾填图题

1760年
1780年
1800年
1820年
1840年

铅笔填涂示范

选择文件（不超过4M）

Fig. 2 The coloring task for the first argumentation activity in the Star of Questionnaires platform. (**a**) Instructions for the coloring item: please use your pencil to depict the gradual change in wing coloring in the five periods of the Industrial Revolution. (**b**) The original version of coloring item. (**c**) Students drew five simple circles to represent the peppered moths and used a pencil to shade the circles to represent the magnitude of change in peppered moth coloring. (**d**) When students completed the coloring task, they took photos of their work and uploaded the photos to the Star of Questionnaires platform

2018. The issue of genetically edited human babies was an effective prompt for argumentation about the pros and cons of the implementation of human gene-editing techniques. Background materials on this topic were prepared for argumentation, including an introductory video about human gene-editing techniques, official population statistics regarding human immunodeficiency virus-positive patients in the past 5 years in China, and public discussion from different communities of people. The starting questions for students to begin argumentation in both the Star of Questionnaires and Collaborative Editing Platform were “Please first share your opinion about the implementation of human gene-editing techniques, then think of the advantages and disadvantages of the implementation of human gene-editing techniques?” and “Please attempt to draw a picture to represent the interrelationships among the community sectors that were relevant to the implementation of human gene-editing techniques from the standpoint of policymakers”. When students completed the pictures, they took photos of their work and uploaded the photos to the Star of Questionnaires platform.

The third topic for argumentation was called *Semmelweis the Obscure*. Ignaz Philipp Semmelweis (1 July 1818–13 August 1865) was a Hungarian physician, now known as the “savior of mothers”. Semmelweis discovered that the incidence of childbed fever could be drastically reduced by the use of hand disinfection in obstetrical clinics and proposed the practice of rigorous hand-washing with chlorinated solution in 1847 while working at the Vienna General Hospital’s First Obstetrical Clinic. However, Semmelweis’ ideas were rejected by the medical community, partly because his research paradigm conflicted with established scientific and medical opinions of the time. In 1865, the increasingly outspoken Semmelweis supposedly suffered a nervous breakdown and was committed to a psychiatric hospital by his colleagues. He died 14 days later after being beaten by the guards. Semmelweis’ practice of washing hands in clinics earned widespread acceptance only years after his death, when Louis Pasteur confirmed the *germ theory*, and Joseph Lister, acting on the French microbiologist’s research, practiced and operated using hygienic methods, with great success.

Semmelweis’ rejected proposal and his tragic death were an effective prompt for argumentation regarding the reasons Semmelweis failed to convince others of the practice of washing hands. Reading materials on this topic had also been prepared for argumentation, including information about Semmelweis’ family and early life, his work on the causes of childbed fever mortality, his efforts to reduce childbed fever, and his breakdown and death (adapted from Archila et al.’s, 2020 study). The prompts for students to begin argumentation in both the Star of Questionnaires platform and Collaborative Editing Platform were “Analyze the materials of historical background and attempt to elaborate the reasons why Semmelweis’ practice of washing hands was rejected” and “Please attempt to draw a picture to represent the interrelationships among the community sectors that were relevant to implementing the practice of washing hands from the standpoint of policymakers”. When students completed the pictures, they took photos of their work and uploaded the photos to the Star of Questionnaires platform.

4.3 Experimental Procedure and Design

Two types of experimental designs were employed in this study to examine our three research questions. First, we used a between-subject experimental design. During the first and second argumentation activities, participants were randomly assigned into one of two conditions (see Fig. 3): collaborative argumentation (experiment group) or individual argumentation (control group). Each student’s levels of conceptual understanding of the first and second topics were assessed with a pre-test, immediate post-test, and delayed post-test. Through comparison with changes in the levels of conceptual understanding among the students in the two conditions across the three assessment phases, the role of collaborative

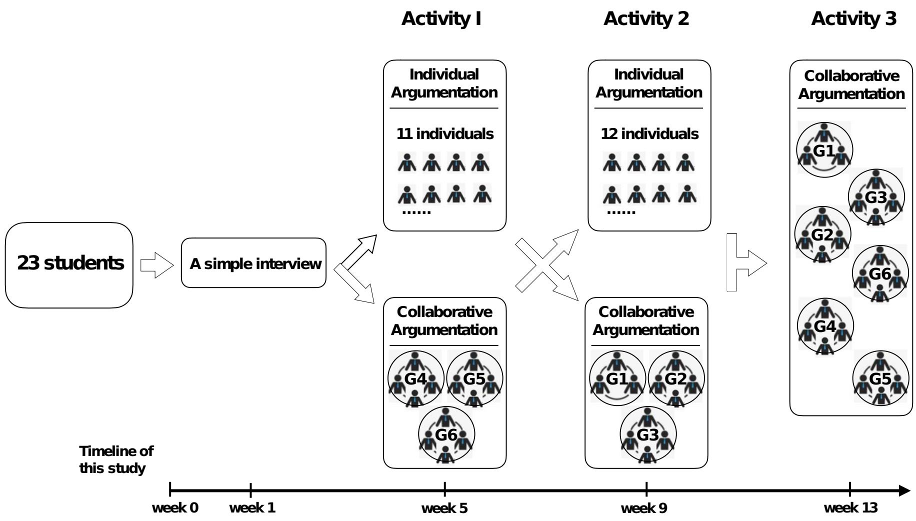

Activity I
Activity 2
Activity 3
Individual Argumentation
11 individuals
......
Individual Argumentation
12 individuals
......
Collaborative Argumentation
G1
G3
G2
G6
G4
G5
23 students
A simple interview
Collaborative Argumentation
G4
G5
G6
Collaborative Argumentation
G1
G2
G3
Timeline of this study
week 0
week 1
week 5
week 9
week 13

Fig. 3 Overview of research design

argumentation on students' conceptual change could be clearly illustrated. The other was within-subject experimental design. The third argumentation activity was conducted with a single group experimental design. During the third argumentation activity, all students participated in the same condition (i.e., collaborative argumentation). Each student's levels of conceptual understanding of the third topic were also assessed in a pre-test, immediate post-test, and delayed post-test. Through analyzing conceptual change between immediate post-test and delayed post-test, high-performing groups and the low-performing groups were distinguished. Differences in types of argumentative dialogue between the high-performing groups and the low-performing groups were further explored.

Before the study began, all students were invited to take a simple interview via Zoom Rooms for 10 min. All students in the simple interview stated that they did not have previous experience of argumentation-based learning in an educational context. Thus, the students could be considered laypersons for argumentation in the context of science education, and students were assumed to have the same level of argumentation skills. Before starting the argumentation activities, each student was requested to complete a series of pre-tests on learners' characteristics on the Star of Questionnaires platform, including motivation, scientific belief, nature of science (NOS), and risk awareness factors of human gene-editing techniques. There were six stages in each argumentation activity (see Fig. 4). From the first to the third argumentation activity, each student's levels of conceptual understanding of the three topics were assessed in a pre-test (stage 1), immediate post-test (stage 4: following collaborative or individual argumentation), and delayed post-test (stage 6: 1 month later). For stage 2 of each argumentation activity, all students first downloaded the background materials on the given topic from the Tencent QQ platform. After reading for 10 min, students were required to complete and submit their individual initial arguments on the Star of Questionnaires platform.

Individual or collaborative argumentation was implemented in stage 3. The arrangements for stage 3 across the three argumentation activities were different. For stage 3 in the first argumentation activity, 23 students were randomly assigned into one of two

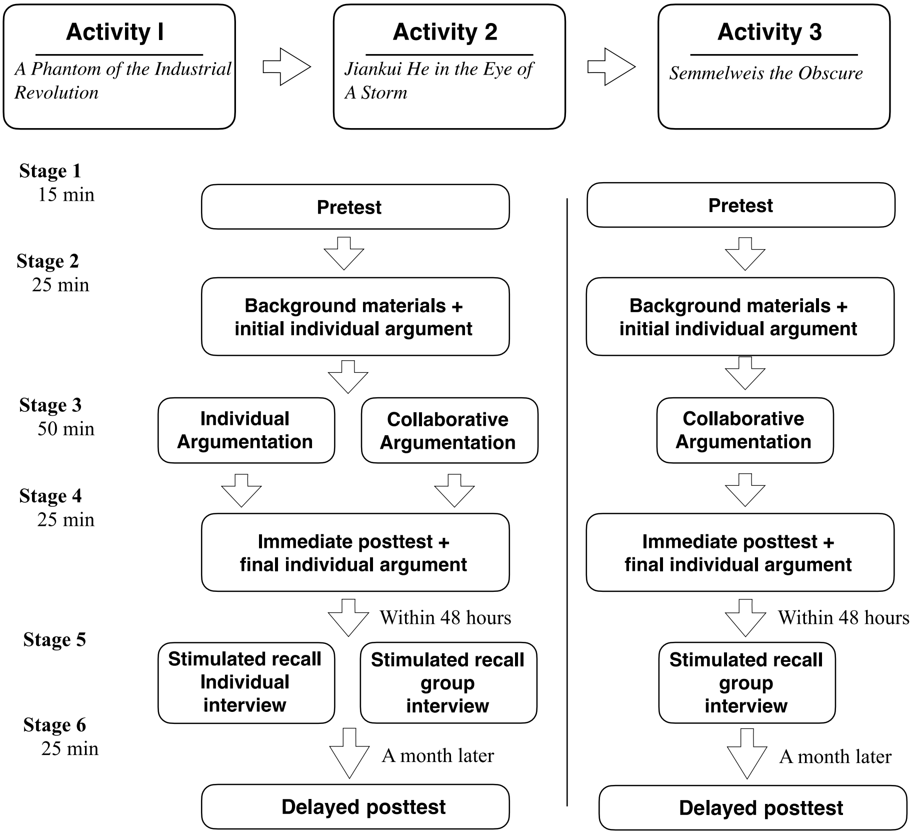

Activity I
A Phantom of the Industrial Revolution
Activity 2
Jiankui He in the Eye of A Storm
Activity 3
Semmelweis the Obscure
Stage 1
15 min
Stage 2
25 min
Stage 3
50 min
Stage 4
25 min
Stage 5
Stage 6
25 min
Pretest
Background materials + initial individual argument
Individual Argumentation
Collaborative Argumentation
Immediate posttest + final individual argument
Within 48 hours
Stimulated recall Individual interview
Stimulated recall group interview
A month later
Delayed posttest

Fig. 4 Procedure of the three argumentation activities

conditions: collaborative argumentation or individual argumentation. There were 12 participants in the collaborative argumentation condition, while the other 11 participants were in the individual argumentation condition regarding the first argumentation activity. The 12 participants in the collaborative argumentation condition were further randomly divided into three groups of four students (groups 4, 5, and 6). Guided by the collaborative argumentation version of the worksheet displayed on the Collaborative Editing Platform, participants from groups 4, 5, and 6 collaboratively carried out the first argumentation activity, while the other 11 participants individually carried out the first argumentation activity in accordance with the online worksheet of the version of individual argumentation.

To maintain layperson status for collaborative argumentation, for stage 3 of the second argumentation activity, the participants in the two conditions were switched to the opposite conditions; thus, the 11 participants who were previously in the individual argumentation condition were switched to the collaborative argumentation condition, and vice versa (see Fig. 4). Therefore, students who participated in the collaborative argumentation condition for the second topic were also laypersons for collaborative argumentation. These 11 participants were further randomly divided into three small groups (groups 1, 2, and 3). There were three students in group 1, and the other two groups had four students. Guided by the collaborative argumentation version of the worksheet displayed on the Collaborative Editing Platform, participants in groups 1, 2, and 3 collaboratively carried out the first argumentation activity, while the other 12 participants individually carried out the first argumentation activity in accordance with the online worksheet of the version of individual argumentation. Regarding stage 3 of the third argumentation activity, there was only one condition, and all participants from the aforementioned six groups participated in collaborative argumentation. By the end of stage 3 of each argumentation activity, participants in the collaborative argumentation condition were required to submit a common argument that represented the product of their collaborative argumentation. The Collaborative Editing Platform provided an online collaborative space enabling participants to type, edit, and elaborate their own and group members' points of view and relevant justifications to construct a common argument within groups, while they verbally discussed the given topics in the online Zoom Rooms platform. Moreover, the discussion and images in each group were recorded with the Zoom Rooms online platform. At stage 4 of each argumentation activity, each participant submitted their final individual arguments on the Star of Questionnaires platform.

Stimulated recall interview was an introspective technique for gaining insight into cognitive processes and implicit beliefs as videotaped passages of dialogue, behavior, and interaction were replayed to participants to stimulate recall of their concurrent cognitive activity and beliefs (Gass & Mackey, 2013; Koltovskaia, 2020). Therefore, after completing each argumentation activity within 48 h (stage 5), students in the collaborative argumentation condition were instructed to engage in group stimulated recall interviews, in which they watched their discussion videos, individuals' original arguments and graphs, groups' common arguments and graphs, and individuals' final arguments and graphs. Participants were asked to recall their thoughts at the time of key events with prompts such as "Did you like this topic?", "What advantages and disadvantages of collaborative argumentation did you see?", "What were your initial thoughts about the group's ideas?", "Did you think it was essential to achieve a consensus within the group? Why?", "What stopped you from further questioning the group when you were unsure of their statements?" "How did you know the information you contributed?", "Which statements made you concede and accept others' arguments?", "Had you thought of alternative explanations, and what made you abandon them?", and "Why did you finally decide to remove these ideas from your individual final argument although you previously accepted them in the construction of common arguments?". In addition, at the same time, an equivalent number of students in the individual argumentation condition were also instructed to complete individual stimulated recall interviews in which they examined their initial individual arguments and final individual arguments. These students were asked to recall their thoughts at the time they experienced cognitive conflicts and refined their final arguments with prompts such as "Did you like this topic?", "What advantages and disadvantages of individual argumentation did you see?", "How did you know the information you used?", "Why did you disagree with ideas from the scientific community?", "Why did you decide to absorb some statements from the scientific community even though you did not agree with their ideas?", and "Had you thought of alternative explanations, and why did you abandon them?". All of the stimulated recall interviews were recorded online for analysis.

4.4 Instruments

4.4.1 Conceptual Understanding Tests

One conceptual understanding test was designed for each topic of argumentation activity. Each conceptual understanding test was conducted at pre-test, immediate post-test, and delayed post-test to assess students' conceptual understanding of scientific concepts at the aforementioned time points. Each conceptual understanding test consisted of three parts: two two-tier items and one open-ended construction item. For each two-tier item, students were presented with short descriptive background material and then judged five true/false statements. Each true/false statement targeted one of the five scientific concepts in theories of the given topic (see Table 2). Participants were further prompted to explain why the statement was correct or incorrect when they gave their judgment. To examine students' level of misconception regarding scientific concepts, approximately half of the statements were incorrect (counterbalanced per scientific concept). In the open-ended construction item, students were required to give a full explanation of how the given natural phenomenon was generated and developed.

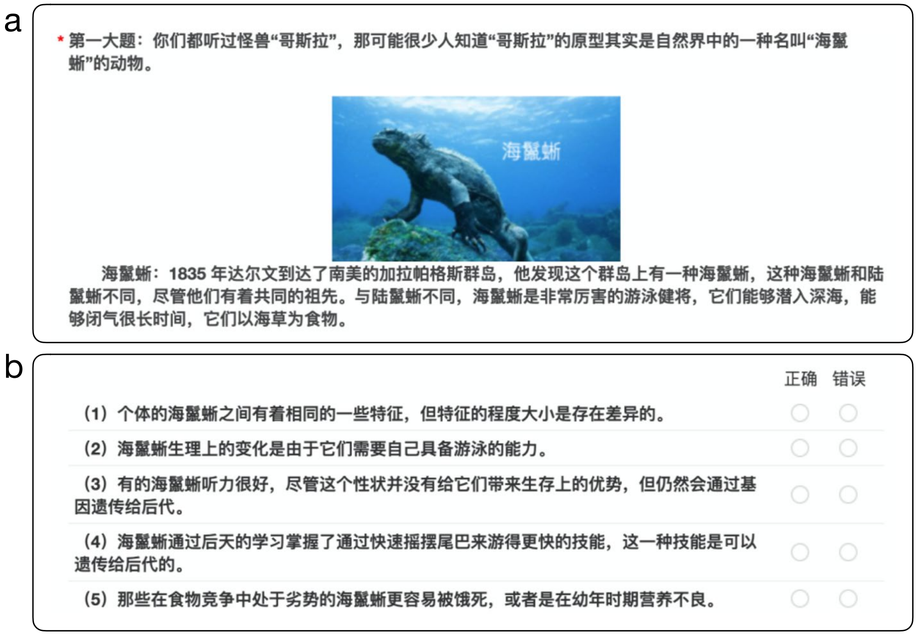
The reason for including two-tier items was based on previous studies reporting that participants often did not refer to each level of a triangular scientific concept in open-ended construction items (Asterhan & Resnick, 2020). Although the overall schema of change that was alluded to could be deduced from these responses, it was difficult or even impossible to assess whether students understood a particular level of the given scientific concept. The combination of two-tier with open-ended construction items allowed for a fine-grained and comprehensive assessment of students' conceptual understanding. Content validity and face validity for each test were established and examined by a panel of three evaluators (one science teacher, one postgraduate biology student, and one professor of clinical medicine domains), ensuring that the tests were relevant to the topics of the three argumentation activities. For the first argumentation activity, internal reliability was good for pre-test (Cronbach's $\alpha = .79$), immediate post-test (Cronbach's $\alpha = .79$), and delayed post-test (Cronbach's $\alpha = .82$). Regarding the second argumentation activity, internal reliability was good for pre-test (Cronbach's $\alpha = .80$), immediate post-test (Cronbach $\alpha = .84$), and delayed post-test (Cronbach's $\alpha = .79$). For the third argumentation activity, internal reliability was good for the pre-test (Cronbach's $\alpha = .81$), immediate post-test (Cronbach's $\alpha =$ .82), and delayed post-test (Cronbach's $\alpha = .80$). All of the tests on conceptual understanding were uploaded and administrated by the teacher on the Star of Questionnaires platform. Students completed all of the tests on computers or mobile phones within the given time. Examples of the conceptual understanding test with two types of item formats were presented on the Star of Questionnaires platform (see Fig. 5 and Fig. 6). And the complete English versions of the conceptual understanding test with two types of item formats are also presented in Appendix 3.

4.4.2 Coding of Conceptual Understanding Tests

Regarding the two-tier items in the conceptual understanding tests, each true/false statement targeted one of the five scientific concepts in theories of the given topic. When coding students' responses to the true/false statements, the indicated choice of right or wrong and the accompanying textual explanation were considered together. A correct choice together with a correct and sufficient explanation resulted in full credit (1). An incorrect true/false choice with an incorrect explanation resulted in zero points. When these two components were not aligned, points were assigned based on the latter explanation. Most of these cases revealed that students had partial conceptual understanding (0.5 points), but there were several cases that showed clear misconceptions in the latter explanations with the correct choice of true/false (0 points).

Table 2 Scientific concepts targeted in conceptual understanding tests for three argumentative activities

	Representation levels	Scientific concepts in argumentation activity 1	Scientific concepts in argumentation activity 2	Scientific concepts in argumentation activity 3
1	Macro	Survival	Phenotype	Reservoir of infection
2	Macro	Learned behaviors	CD4 and CCR5 proteins	Susceptible population
3	Symbolic	Intra-species diversity	Virus transfection	Route of transmission
4	Symbolic	Proportional and typological change	Virus attachment	Maternal mortality rate
5	Sub-micro	Inheritance of traits	Clustered regularly interspaced short palindromic repeats gene editing	Pathogenic microorganism

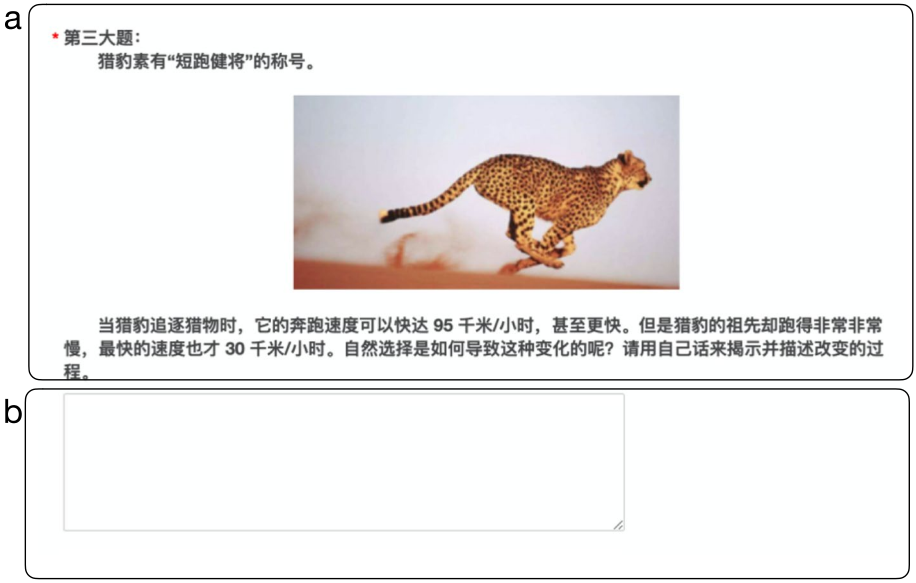


* 第一大题：你们都听过怪兽“哥斯拉”，那可能很少人知道“哥斯拉”的原型其实是自然界中的一种名叫“海鬣蜥”的动物。

海鬣蜥：1835 年达尔文到达了南美的加拉帕格斯群岛，他发现这个群岛上有一种海鬣蜥，这种海鬣蜥和陆鬣蜥不同，尽管他们有着共同的祖先。与陆鬣蜥不同，海鬣蜥是非常厉害的游泳健将，它们能够潜入深海，能够闭气很长时间，它们以海草为食物。

	正确	错误
（1）个体的海鬣蜥之间有着相同的一些特征，但特征的程度大小是存在差异的。	○	○
（2）海鬣蜥生理上的变化是由于它们需要自己具备游泳的能力。	○	○
（3）有的海鬣蜥听力很好，尽管这个性状并没有给它们带来生存上的优势，但仍然会通过基因遗传给后代。	○	○
（4）海鬣蜥通过后天的学习掌握了通过快速摇摆尾巴来游得更快的技能，这一种技能是可以遗传给后代的。	○	○
（5）那些在食物竞争中处于劣势的海鬣蜥更容易被饿死，或者是在幼年时期营养不良。	○	○

Fig. 5 Example of a two-tier item in a conceptual understanding test. (**a**) A short description of a species; (**b**) five true/false statements that targeted five scientific concepts from theories related to the given topic. Students were instructed to judge each statement and further explain why it was correct or incorrect



* 第三大题：

猎豹素有“短跑健将”的称号。

当猎豹追逐猎物时，它的奔跑速度可以快达 95 千米/小时，甚至更快。但是猎豹的祖先却跑得非常非常慢，最快的速度也才 30 千米/小时。自然选择是如何导致这种变化的呢？请用自己话来揭示并描述改变的过程。

Fig. 6 Example of an open-ended construction item on a conceptual understanding test. (**a**) A short description of a natural phenomenon that occurred in a species; (**b**) students were instructed to give a full explanation of how the given natural phenomenon was generated and developed

Regarding the open-ended construction items in the conceptual understanding tests, coding of students' responses was based on the overall model of theories in the given issues, rather than only the particular scientific concept that was targeted. Solutions that contained no misconceptions and correctly explained how a given phenomenon was generated and developed received full credit (1). Answers that were partially correct or contained both correct as well as incorrect aspects received .5 points. For omissions, misconceptions, or other crucial errors, answers were given 0 points. This coding procedure was adapted from the coding procedure developed in Asterhan and Dotan's (2018) study.

Following a training period of 4 h, two human coders randomly selected and coded 35 students' responses in the conceptual understanding tests (16.91% of the total data set). Inter-rater reliability was satisfactory, $0.72 < \text{Cohen's } k < 0.79$. Differences were resolved through discussion, after which the entire data set was coded. The total score on each conceptual understanding test was compiled by adding the different weight for each test item, while assigning the open-ended construction item score a weight of 5 points (instead of 1). Each student's total score was then transformed into percentage scores ranging from 0 to 100.

To examine the value of collaborative argumentation for conceptual change, we classified the students according to the conditions of argumentation for the first and second argumentation activities, respectively (i.e., students in the individual argumentation condition or collaborative argumentation condition). We then compared conceptual understanding scores among the two groups at the three phases of pre-test, immediate post-test, and delayed post-test for the first and second argumentation activities. Differences in conceptual gains for each activity from pre-test to delayed post-test, from pre-test to immediate post-test, and from immediate post-test to delayed post-test were then analyzed and compared among the two groups to examine whether students in the two conditions differed in conceptual change during the aforementioned time periods. Thus, the effects of conditions of argumentation on concept change, particularly substantive conceptual gains (i.e., conceptual gains from immediate post-test to delayed post-test; adapted from Asterhan and Resnick (2020)) were further elaborated.

The mean normalized scores for conceptual understanding tests were calculated per condition across the three phases of the first and second argumentation activity. Distributions of scores were checked for outliers. Score residues were checked for normality assumptions by inspecting skewness and kurtosis (< 1) and the Kolmogorov-Smirnov test of normality. The Kolmogorov-Smirnov test indicated that the assumption of normality was not violated for the normality of scores in the first and second argumentation activities ($p > .80$).

Mean normalized gain scores were calculated per condition from pre-test to delayed post-test, from pre-test to immediate post-test, and from immediate post-test to delayed post-test in the first and second argumentation activities. Distributions of gain scores were checked for outliers. Gain score residues were checked for normality assumptions by inspecting skewness and kurtosis (< 1) and the Kolmogorov-Smirnov test of normality. The Kolmogorov-Smirnov test results indicated that the assumption of normality of gain scores was not violated for the first and second argumentation activities ($p > .80$).

Based on the results of substantive concept gains, we further attempted to explore the associations between conceptual change and argumentative dialogue for the first and second argumentation activities. According to the mean substantive conceptual gains at the group level, the top one-third of the ranking was classified as the high-performing group, and the bottom one-third was classified as the low-performing group. There were three collaborative groups for the first and second argumentation activities. Accordingly, one high-performing group and one low-performing group were identified in each argumentation activity. We then examined whether there were differences in argumentative dialogue between the high-performing group and the low-performing group in the first and second argumentation activities. All data processing and statistical analyses were performed using SPSS Statistics 25.0.

4.4.3 Coding of Dialogue

First, 12 video-recorded discussions were transcribed. The mean length of these video-recorded discussions was 49:98 min (ranging from 45:04 to 58.81 min). Transcriptions included all verbal content, intonation, other auditory features, and brief descriptions on a screencast projected by students within groups. Second, discussions in the four groups were segmented into episodes as the unit of dialogue analysis. The segmentation of the episodes was performed by two coders independently, and the agreement of the segmentation was .92, indicating an appropriate level of reliability. After the two coders resolved disagreements in segmentation, the rest of the discussions were segmented (Chang et al., 2017). Finally, following a top-down scheme proposed by Asterhan and Schwarz (2016), initial coding efforts focused on three types of argumentative dialogue, considering whether the discussion could be characterized as deliberative argumentation, disputative argumentation, and co-consensual construction. The coding category, definitions, and examples are shown in Table 3. Recordings of dialogue were transcribed and coded by two coders using NVivo software (Version 12, QRS International). The two coders independently coded the four discussions that had been segmented. Inter-rater agreement was satisfactory, and Cohen's *k* ranged from .82 to .85.

To confirm the association between conceptual change and argumentative dialogue found in the first and second argumentation activities, we further extended the sample of students in the collaborative argumentation condition and re-examined whether the association between conceptual change and argumentative dialogue also existed in the third argumentation activity. All six groups of students participated in the third argumentation activity. According to the mean substantive conceptual gains at the group level, the top third of the ranking was classified as the high-performing group, and the bottom third was classified as the low-performing group. We then compared argumentative dialogue between the two groups for the third argumentation activity. All data processing and statistical analyses were performed using SPSS Statistics 25.0.

4.4.4 Coding of Stimulated Recall Interview Reponses

Thirty-five videos of stimulated recall interviews included 12 group interviews and 23 individual interviews. The 12 group interviews included three group interviews for the first argumentation activity, three group interviews for the second argumentation activity, and six group interviews for the third argumentation activity. Twenty-three individual interviews included 11 individual interviews for the first argumentation activity and 12 individual interviews for the second argumentation activity. The mean length of group interviews videos was 90:33 min (ranging from 89:04 to 98.01 min). The mean length of individual interview videos was 10:50 min (ranging from 8:06 to 18.55 min).

Recordings of interviews were transcribed and coded by two coders for themes using NVivo software (Version 12, QRS International). First, the initial coding was

Table 3 Coding category for argumentative dialogue

Coding category of argumentative dialogue	Definition and examples
Deliberative argumentation	A discussion focused on maintaining a balance between critical reasoning and collaborative knowledge construction. This dialogue is not *position*-driven but *issue*-driven. The dialogue between discussants shows a willingness to make concessions in response to sound or cogent arguments. ***Examples:*** Student A: If the wing color is controlled by a pair of alleles: white wing is the dominant phenotype while black wing is the recessive phenotype. Then, uh, several dark-colored peppered moths should exist among the population of peppered moths even before the start of the Industrial Revolution. Student B: Oh, do you mean that the pair of alleles on wing color provide the raw material on which evolutionary forces such as natural selection can act? Student A: Yes. Student B: The question is, uh, whether that process happens suddenly or gradually? What's your opinion? Student A: I think it might happen gradually. Student B: [point at the background materials]. It shows the population data of the peppered moths in the five periods of the Industrial Revolution. So, yes, you are right, dark-colored peppered moths are well adapted to the new environments, and they survive and continue themselves from generation to generation. That process needs a long period. Student A: Such explanations are reasonable. I totally understand how it evolves.
Disputative argumentation	A discussion in which discussants defended a viewpoint and undermined alternative viewpoints to convince opponents to switch sides. This dialogue type focuses not on the collaborative co-construction of knowledge but on the interpersonal competitive dimension of social interaction. ***Examples:*** Student A: You two make me confused. The source of this mutation … (interrupted by student B) Student B: Oh, you mean that the light-colored peppered moths become blackened by soot, because the soot is a kind of irritating chemical colorant. Student C: No, that is impossible! The black color would have always existed in genes of the light-colored peppered moths. [short silence] Student B: I don't think so. I believe the light-colored peppered moths are colored by a certain external substance, such as soot. Student C: No! That is not a coloring process. I don't think this mutation is induced by soot. It feels like it came out of nowhere.
Consensual co-construction	A discussion in which discussants transacted on each other's contributions by agreeing with, elaborating, expanding ideas. This dialogue between discussants shows development and consolidation of a one-sided argument, rather than challenging or juxtaposing different alternatives. ***Examples:*** Student A: I think Semmelweis needed to conduct a between-subjects experimental design to explore the relationship between hand-washing and child fever. Student B: Hum Student A: And if the incidence of child fever in the treatment group decreases then it could confirm the effectiveness of the practice of hand-washing on child fever. Student B: Yes. Student C: Yes. Student A: It also needs to control for irrelevant variables such as the weight of mothers and the procedure of delivering children. Student C: Yes. That sounds reasonable.

open (Corbin, 1990), and students' responses were coded by category as they arose. As coding progressed, it became clear that a number of categories of responses were present (Bryman, 2016). For example, several students in the collaborative argumentation condition mentioned that group members' misconceived statements triggered them to reflect upon whether there were flaws in their own ideas, and this was coded as "Others' misconceptions triggered self-reflection". Thus, "Others' misconceptions triggered self-reflection" was classified as a subtheme. Second, interviews were thematically coded three times to consistently establish categories (i.e., subthemes). Because no more categories arose, it can be assumed that theoretical saturation had been reached (Bryman, 2016). Third, we classified all of the subthemes into three themes: cognitive aspects, ontological aspects, and intentional aspects. In addition, if some subthemes could not be classified into the three themes, two coders would be assisted with four science educators to further scrutinize and discuss these subthemes to label them. Finally, we counted and added together the frequency of subthemes in coding of each student's response, as well as the relevant number of students.

To elucidate the opportunities collaborative argumentation could offer for changes in cognitive, ontological, and intentional aspects, we then counted and added together with the frequency of subthemes in coding of each student's response to stimulated recall interviews as well as the relevant number of students. Opportunities provided by the two conditions were then compared and used to clarify the reasons that collaborative argumentation could or could not promote conceptual change.

Besides, following the previous studies (Breitmayer, 1991; Carter et al., 2014; Rich, 2009), several techniques were employed in order to establish trustworthiness of the data collection, analysis interpretation, and reporting, including peer debriefing, data source triangulation, and member checks. A peer debriefing was completed by having a professor with a formal education in qualitative methods (a minimum of 3 qualitative research methodology courses at the doctoral level) review the documented coding categories of argumentative dialogue and coding categories of stimulated recall interview responses for relevance, consistency, and logic. Moreover, the reviewer examined the stimulated recall interview questions in each transcript to determine if they were "leading" in nature. The textual data from any questions identified as being leading were not included in the analysis. The reviewer agreed with the findings based on the purpose of the study. The second technique was data source triangulation. With the cross-checking perspectives, we used multiple data sources with similar foci to obtain diverse views about a topic or the purpose of validation. For example, after completing each argumentation activity within 48 h, students in the collaborative argumentation condition were instructed to engage in group stimulated recall interviews; at the same time, an equivalent number of students in the individual argumentation condition were also instructed to complete individual stimulated recall interviews. The third technique was member checks, which were completed electronically by e-mailing the results to 8 participants and allowing them to comment on the coding categories. Five individuals responded, agreed with the results, and had no further input, indicating no misinterpretation of the argumentative dialogue and the stimulated recall interviews that emerged from this study. On an informal basis, I also explained the results to 4 other participants, and they agreed with the findings.

5 Results

In this section, we presented the results according to each research question.

5.1 Results About Research Question 1 "Do Students who Engage in Collaborative Argumentation Demonstrate Better Conceptual Change Compared with Students who Engage in Individual Argumentation?"

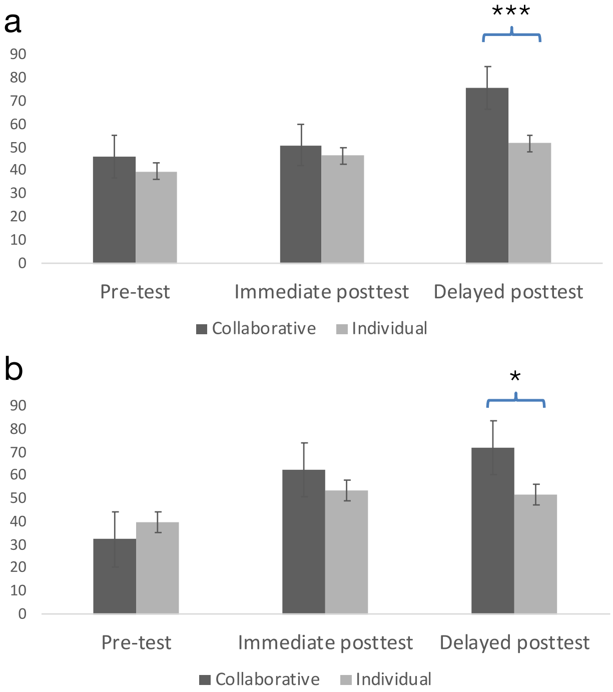
5.1.1 Effects of Collaborative Argumentation on Conceptual Understanding

To examine the effects of collaborative argumentation on conceptual understanding, mean scores in all three phases of the first and second argumentation activity were compared between students in the collaborative argumentation condition and students in the individual argumentation condition. Comparisons were conducted with paired sample *t* tests. The results revealed students in the collaborative argumentation condition exhibited better conceptual understanding in the delayed phase of the two argumentation activities compared with students in the individual argumentation condition (see Fig. 7). Regarding the first argumentation activity, students in the collaborative argumentation condition exhibited significantly better conceptual understanding ($M = 75.92$, SD = 10.65, $N = 12$) in the delayed post-test phase compared with students in the individual argumentation condition ($M = 51.91$, SD = 22.81, $N = 11$, t [21] = 3.16, $p < .001$), while there were not statistically significant differences in conceptual understanding in the pre-test and immediate post-test phases (see Fig. 7a). Regarding the second argumentation activity, students in the collaborative argumentation condition exhibited better conceptual understanding ($M = 77.27$, SD = 14.21, $N = 11$) in the delayed post-test phase compared with students in the individual argumentation condition (M = 52.08, SD = 21.06, N = 12), t (21) = 2.64, $p < .05$, while there were not statistically significant differences in conceptual understanding in the pre-test and immediate post-test phases (see Fig. 7b).

5.1.2 Effects of Collaborative Argumentation on Conceptual Change

To examine the effects of collaborative argumentation on conceptual change, mean gain scores from pre-test to delayed post-test, from pre-test to immediate post-test, and from immediate post-test to delayed post-test in the first and second argumentation activity were compared between students in the collaborative argumentation condition and students in the individual argumentation condition. Comparisons were conducted with paired sample *t* tests. Regarding the first argumentation activity, students in the collaborative argumentation condition showed significantly greater gains from pre-test to delayed post-test ($M = 22.92$, SD = 10.65, $N = 12$) compared with students in the individual argumentation condition ($M = 5.27$, SD = 15.11, $N = 11$), t (21) = 3.16, $p <$ .001. Students in the collaborative argumentation condition showed significantly greater gains from immediate post-test to delayed post-test ($M = 24.67$, SD = 13.51, $N = 12$) compared with students in the individual argumentation condition ($M = 1.09$, SD = 22.81, $N = 11$), t (21) = 3.17, $p < .001$. There were not statistically significant differences in conceptual gains from pre-test to immediate post-test between students in the two conditions (see Fig. 8a). Regarding the second argumentation activity, students in the collaborative argumentation condition showed greater gains in the three phases compared with students in the individual argumentation condition. Specifically, students in

Fig. 7 Comparison of conceptual understanding between students in the collaborative argumentation condition and students in the individual argumentation condition for the first and second argumentation activities. (**a**) The first argumentation activity. (**b**) The second argumentation activity. *indicates $p < .05$, **indicates $p < .01$, ***indicates $p < .001$

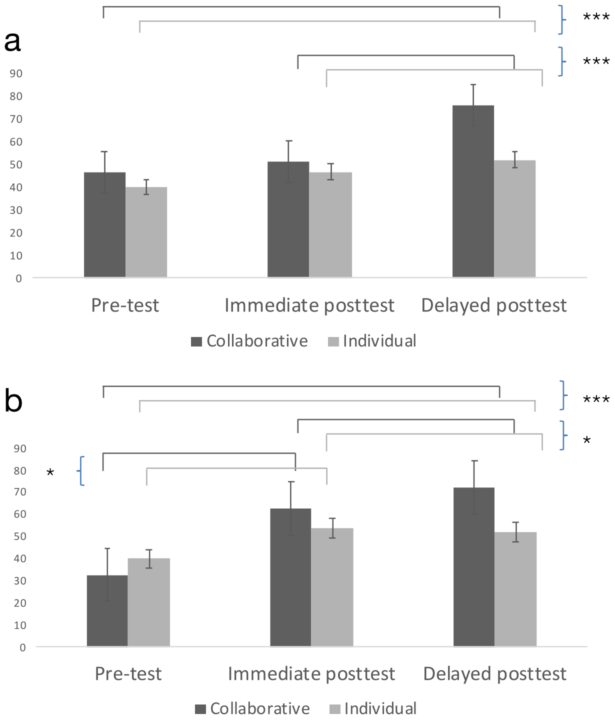


the collaborative argumentation condition showed significantly greater gains from pre-test to delayed post-test ($M = 39.55$, SD = 12.54, $N = 11$) compared with students in the individual argumentation condition ($M = 12.08$, SD = 16.77, $N = 12$), $t\ (21) = 4.35$, $p < .001$. Students in the collaborative argumentation condition showed greater gains from immediate post-test to delayed post-test ($M = 9.55$, SD = 7.57, $N = 11$) compared with students in the individual argumentation condition ($M = -1.67$, SD = 12.80, $N = 12$), $t\ (21) = 2.50$, $p < .05$. In addition, students in the collaborative argumentation condition also showed greater gains from pre-test to immediate post-test ($M = 30.00$, SD = 15.90, $N = 12$) compared with students in the individual argumentation condition ($M = 13.75$, SD = 17.34, $N = 12$), $t\ (21) = 2.29$, $p < .05$ (see Fig. 8b).

5.2 Results About Research Question 2 "What Opportunities could Collaborative Argumentation Offer for Change in Cognitive, Ontological and Intentional Aspects of Learning?"

To clarify the opportunities collaborative argumentation can offer for conceptual change, we coded students' responses to stimulated recall interviews, resulting in a number of subthemes. Most of the subthemes were classified into three themes: cognitive aspects, ontological aspects, and intentional aspects. Since some subthemes could not be classified into the three themes, two coders assisted with four science educators to further scrutinize and discuss these subthemes to label them as epistemological aspects. To maintain an equivalent number of participants, the results of coding of stimulated recall interview responses for the first two argumentation activities are presented in Table 4, together with the number

Fig. 8 Comparison of gain scores (conceptual change) between students in the collaborative argumentation condition and students in the individual argumentation condition for the first and second argumentation activities. (**a**) The first argumentation activity; (**b**) the second argumentation activity. *indicates $p < .05$, **indicates $p < .01$, ***indicates $p < .001$



of students who described these subthemes and the frequency with which these subthemes emerged in the two conditions.

The results revealed that, compared with individual argumentation, collaborative argumentation provided more opportunities for facilitating change in cognitive aspects. For example, collaborative argumentation provided opportunities for students to listen to others' misconceptions and accurate understandings, prompting students to self-reflect and perceive the flaws in their own incomplete ideas. In addition, collaborative argumentation also provided opportunities for students to be trapped in a "tug-of-war" situation, prompting students to re-examine whether key evidence was neglected before. This experience provided opportunities for students to juxtapose various opinions and evidence with different levels of validity, which encouraged them to construct and develop epistemological criteria for knowledge evaluation. Thus, the aforementioned opportunities provided by collaborative argumentation resulted in more accurate understanding of scientific concepts, a greater ability to construct sound and cogent arguments, and a greater ability to evaluate knowledge.

Collaborative argumentation provided more opportunities for facilitating change in ontological aspects. For example, it provided opportunities for students to be exposed to others' conceptualization processes of scientific concepts, prompting students to self-reflect and identify flaws in their ways of thinking. In addition, it provided opportunities for students to point out flaws in the logic of group members' arguments, which stimulated them to develop their own logical thinking. Moreover, collaborative argumentation provided opportunities for students to discuss whether experimental design could support the goals of study, and whether evidence could verify hypotheses, scaffolding

Table 4 Categorization of students' responses to stimulated recall interviews

Aspect that aided understanding of scientific concepts	Collaborative argumentation (n = 23)		Individual argumentation (n = 23)	
	Frequency	No. of students	Frequency	No. of students
Cognitive aspects	**88**	**18**	**54**	**13**
Constructing complex evidence-based arguments in final individual arguments	25	16	15	11
Others' misconceptions triggered self-reflection	20	14	10	7
Re-examining evidence to discover details that were previously neglected	15	11	6	6
Adopting others' cogent and sound arguments	13	10	3	2
Acquiring the ability to speak and write the langue of science	11	9	6	4
Developing criteria for knowledge evaluation	4	4	—	—
Not discovering new usable information, and still focusing on known knowledge	—	—	14	13
Ontological aspects	**25**	**9**	**17**	**8**
Others' conceptualization processes inspired and changed their original ways of thinking	13	8	3	3
Developing logical and critical ways of thinking	12	8	3	2
Not perceiving differences between argumentation and traditional teaching approaches	—	—	11	7
Epistemological aspects	**49**	**11**	**17**	**7**
Using sophisticated metacognitive strategies	12	11	8	6
Gaining an overview of the enterprise of science	10	7	2	2
Understanding control variables and experimental design	9	5	—	—
Taking multiple positions to solve problems	6	4	7	4
Requiring more time to argue	12	8	—	—
Intentional aspects	**47**	**12**	**17**	**6**
Increased positive attitudes toward science learning	14	11	—	—
Continuing to think about and discuss topics after class	9	4	—	—
Improving self-confidence about individual arguments	8	7	5	3
Being enthusiastic about discussing science topics	7	4	5	5
Showing a willingness to apply argumentation to their own classes	6	4	4	4
Adding "being a scientist" to a list of goals for the future	3	3	—	—

Table 4 (continued)

Aspect that aided understanding of scientific concepts	Collaborative argumentation ($n = 23$)		Individual argumentation ($n = 23$)	
	Frequency	No. of students	Frequency	No. of students
Feeling depressed about science reading	—	—	3	3

students' knowledge to gain an overview of the enterprise of science and paradigms of scientific investigation.

Collaborative argumentation also provided more opportunities for facilitating radical change in epistemological aspects. For example, collaborative argumentation provided opportunities for students to discuss whether experimental design could support the goals of study, and whether evidence could verify hypotheses, scaffolding students' knowledge to gain an overview of the enterprise of science and paradigms of scientific investigation.

Collaborative argumentation provided opportunities for facilitating change in intentional aspects. For example, it provided opportunities for students to be inspired by others' motivations and beliefs, prompting them to develop positive attitudes toward science learning. In addition, it provided opportunities for students to discuss and think about scientific concepts from macro, symbolic, and sub-micro levels, which relieved students' sense of anxiety about science learning and improved their self-confidence regarding individual explanations. Moreover, because these students already had a sense of identity as pre-service teachers, collaborative argumentation provided opportunities for them to experience the procedures involved in scientific research, which not only motivated some of them to add "being a scientist" to a list of goals for the future, but also inspired willingness in some students to apply argumentation in their future classes.

The quotations that exemplified what opportunities could collaborative argumentation offer for change in cognitive, ontological, epistemological, and intentional aspects of learning were as follows, respectively:

Student Wang: "when my partners discussed whether light-colored pepper moths of the same generation had the same colored wings—some may have more black spots while others may have fewer black spots. I suddenly realized although a generation of light-colored peppered moths shared some common characteristics, the degree of the common characteristics was different".

Student Lin: "I was originally a person with poor critical and logical thinking. And I was always lost in a lot of information. Thanks to my partners, they guided me to sort out the details of how virion attach and entry cells. I think that is very helpful for critical and logical thinking".

Student Chen: "I am always curious about how scientists discover the natural world? and what strategies they should rely on to solve puzzles? After discussing with Li and Sun (two of the group members), I have got some ideas about how control variables are controlled in experiment designs".

Student Zhang: "Although I'm a postgraduate of science education, I never found science is interesting. Until this time—my group members and I debated for a scientific topic, I came to realize I may be a scientist material".

5.3 Results About Research Question 3 "Is Long-Lasting Conceptual Change Associated with a Particular Pattern of Argumentative Dialogue in Collaborative Argumentation?"

To further explore the association between conceptual change and argumentative dialogue, we examined whether students in the high- and low-performing groups differed in argumentative dialogue for the first and second argumentation activities. Comparisons were conducted using Chi-square tests. Regarding the first argumentation activity, there were three groups in the collaborative argumentation condition. Group 5 was a low-performing group, and group 4 was a high-performing group, according to the mean substantive conceptual gains (i.e., gains from immediate post-test to delayed post-test) at the group level. The results revealed that students in the high- and low-performing groups differed in argumentative dialogue (χ^2 [1, 2] = 6.973, $p < .05$). Students in the high-performing group (group 4) exhibited a U-shaped pattern of argumentative dialogue during collaborative argumentation, exhibiting more deliberative argumentation than disputative argumentation, and more co-consensual construction than disputative argumentation (see Fig. 9a).

Regarding the second argumentation activity, there were three groups in the collaborative argumentation condition. Group 2 was a low-performing group, and group 1 was a high-performing group according to the mean substantive concept gains at the group level. The results indicated that students in the high- and low-performing groups differed in argumentative dialogue (χ^2 [1, 2] = 6.202, $p < .05$). Interestingly, students in the high-performing group (group 1) also exhibited a U-shaped pattern of argumentative dialogue during collaborative argumentation, exhibiting more deliberative argumentation than disputative argumentation, and more co-consensual construction than disputative argumentation (see Fig. 9b).

To confirm the association between conceptual change and argumentative dialogue observed in the first and second argumentation activity, we extended the sample of students in the collaborative argumentation condition and retested whether students in the high- and low-performing groups differed in argumentative dialogue for the third argumentation activities. Comparisons were conducted using chi-square tests. There were six groups in the collaborative argumentation condition for the third argumentation activity. Groups 1 and 4 were low-performing groups, and groups 2 and 3 were high-performing groups, according to the mean substantive concept gains at the group level. The results indicated that students in the high- and low-performing groups differed in argumentative dialogue (χ^2 [1, 2] = 14.962, $p = .001$). Students in the high-performing group (groups 2 and 3) also exhibited a U-shaped pattern of argumentative dialogue during collaborative argumentation, exhibiting more deliberative argumentation than disputative argumentation and more co-consensual construction than disputative argumentation (see Fig. 10).

In summary, results of our studies revealed that collaborative argumentation had a delayed but long-lasting effect on conceptual change in science education. Collaborative argumentation provided opportunities for change in cognitive, ontological, and intentional aspects of learning. Dialogue protocol analysis revealed that long-lasting conceptual change was associated with a U-shaped pattern of argumentative dialogue (i.e., more deliberative argumentation than disputative argumentation, and more co-consensual construction than disputative argumentation) in collaborative argumentation.

6 Discussion and Implications

The current findings revealed that collaborative argumentation significantly induced greater conceptual change among students compared with individual argumentation. This advantage was exhibited in a delayed post-test administered 1 month later, rather than immediately following argumentation activities, suggesting a delayed but long-lasting effect on conceptual change in science education. Thus, the effects of change in a short-term variable (i.e., from traditional teaching approaches to collaborative argumentation) did not

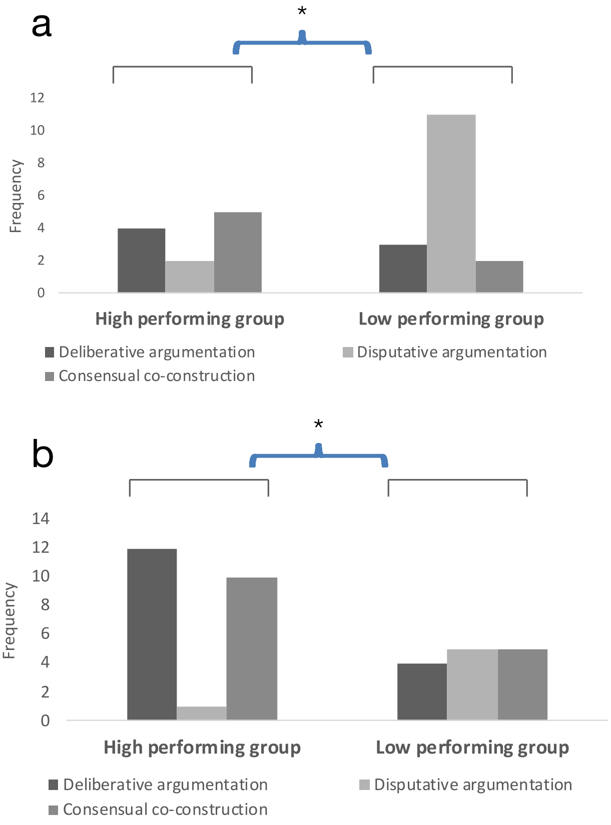


Fig. 9 Comparing argumentative dialogue between the high- and low-performing groups for the first and second argumentation activities. (**a**) For the first argumentation activity, the high- and low-performing groups differed in argumentative dialogue (χ^2 [1, 2] = 6.973, $p < .05$), and the high-performing group exhibited a U-shaped pattern of argumentative dialogue; (**b**) for the second argumentative activity, high- and low-performing groups differed in argumentative dialogue (χ^2 [1, 2] = 6.202, $p < .05$), and the high-performing group exhibited a U-shaped pattern of argumentative dialogue. *indicates $p < .05$, **indicates $p < .01$, ***indicates $p < .001$

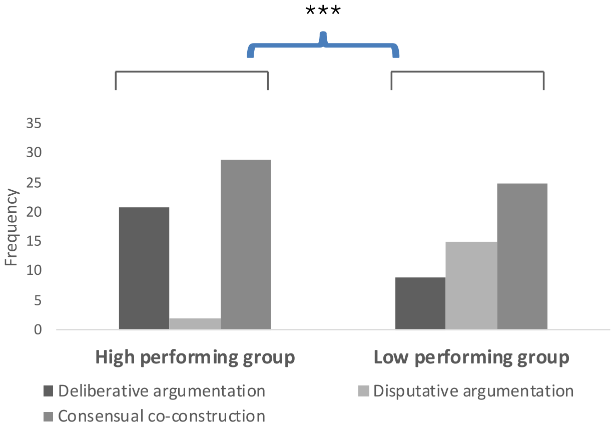


Fig. 10 Comparing argumentative dialogue between the high- and low-performing groups for the third argumentation activities. High- and low-performing groups differed in argumentative dialogue χ^2 (1, 2) = 14.962, $p = .001$, and the high-performing group exhibited a U-shaped pattern of argumentative dialogue. *indicates $p < .05$, **indicates $p < .01$, ***indicates $p < .001$

reflect an immediate representation at the moment of teaching but had a greater impact on long-term performance (i.e., conceptual change) after a delay period (Hall et al., 1984; Sampson & Clark, 2009, 2011). Why did collaborative argumentation show a delayed effect on conceptual change? Some scholars of conceptual change previously reported that many misconceptions were counter-intuitive and structured with triangular levels, showing robustness and resistance to teaching approaches and educational interventions (Anderson & Smith, 1987; Chi, 2008; Vosniadou & Mason, 2012). Thus, although a teaching approach or educational intervention could facilitate change from misconception to accurate understanding of scientific concepts, it may also require more time to explicitly show its effectiveness. Two previous studies used pre-test and post-test measures of conceptual understanding or scientific concept assessments, reporting no effects of collaborative argumentation on conceptual change in the short period (Asterhan & Resnick, 2020; Liu et al., 2019). Findings in this study not only echoes the previous studies but also further extends our understanding of the value of collaborative argumentation for conceptual change—the value of collaborative argumentation for conceptual change should be investigated and discussed under the broadened view of long-lasting science learning (Asterhan & Resnick, 2020; McLure et al., 2020; Sampson & Clark, 2009, 2011).

The second research question concerned the opportunities that collaborative argumentation could offer for conceptual change. Compared with individual argumentation, collaborative argumentation provides more opportunities for facilitating change not only in cognitive aspects, but also in ontological, epistemological, and intentional aspects. This finding is in accord with previous proposals for the value of taking multiple positions to understand and support conceptual change in science education (Duit & Treagust, 2012; McLure et al., 2020). Teaching approaches that take a single epistemological position to promote conceptual change have been reported to result in limited effects, such as fragmentary knowledge revision and short-lived conceptual change (Duit & Treagust, 2012; McLure et al., 2020). In contrast, the highly open environments provided by collaborative argumentation have been reported to facilitate change in the cognitive, ontological, and intentional aspects, resulting in long-lasting conceptual change and, more importantly, resilient science learning (McLure et al., 2020; Sinatra, 2003; Sinatra & Pintrich, 2003; Vosniadou, 2003).

Finally, we further explored whether long-lasting conceptual change was associated with a particular pattern of argumentative dialogue in collaborative argumentation. The results confirmed that long-lasting conceptual change was associated with a U-shaped pattern of argumentative dialogue in collaborative argumentation. This unexpected finding not only supports the suggestions of previous studies on collaborative argumentation considering the roles of verbal dialogue (Asterhan & Resnick, 2020; Liu et al., 2019; Yang et al., 2015), but also has evidenced that conceptual gains from argumentation activities were contingent on productive verbal dialogue powered by collaborative argumentation. These findings raise the question of what productive dialogue in collaborative argumentation looks like. Regarding this question, deliberative argumentation has been described by many scholars of argumentation as an idealized form of argumentative dialogue (Asterhan & Schwarz, 2016; Berland & Hammer, 2012; Felton et al., 2009). For example, Asterhan and Babichenko's (2015) findings indicated that, in tightly controlled and scripted settings, better conceptual understanding at post-test was associated with deliberative argumentation when individuals engaged in interactions with a human or a virtual peer who used deliberative or disputative dialogue. In contrast, the current study was conducted in an actual science classroom, revealing that long-lasting conceptual change was associated with a U-shaped pattern of argumentative dialogue, rather than one particular type of argumentative dialogue. This finding elucidates the complexity of the roles of argumentative dialogue in conceptual change in the context of science education (Asterhan & Schwarz, 2016). Therefore, the current findings extend present research by demonstrating different types of productive verbal dialogue in the different contexts of collaborative argumentation.

The criteria of productive verbal dialogue in collaborative argumentation should consider this context-dependence, because collaborative learning was context-dependent (Asterhan & Schwarz, 2016; Fransen et al., 2011). As a potential explanation for this finding, the three distinct types of argumentative dialogue in the context of collaborative argumentation may have provided different opportunities for science learning (Asterhan & Resnick, 2020; Asterhan & Schwarz, 2016). Thus, deliberative argumentation, disputative argumentation, and consensual co-construction might supplement each other to promote conceptual change in students. For example, in groups of four students, when a student quarreled with another student and lost the argument, the two other students advised and guided them to calm down and to return to the main topic of the activity. These advising and guiding processes included statements such as “I think we should analyze and compare the advantages and disadvantages of each person’s idea” and “I am in favor of your idea because I also thought so at the beginning”. The former was directed to deliberative argumentation, providing opportunities for group members to apply rational thinking to reexamine ideas and relevant evidence to restructure more cogent common arguments. The latter was directed to consensual co-construction, providing opportunities for group members to consolidate one-sided arguments, which may alleviate the “losing” student’s unpleasant experience, and also generate a sense of group belonging among the whole group.

Overall, the current study not only confirmed the value of collaborative argumentation for conceptual change in science education but also revealed an association between long-lasting conceptual change and argumentative dialogue induced by collaborative argumentation. Therefore, implications for conceptual change in science education could be further elucidated on theoretical, methodological, and practical aspects. Regarding theoretical aspects, the implication is that the perspective of conceptual change in science education should be broadened from merely focusing on cognitive dimension to considering ontological and intentional dimensions. Regarding methodological aspects, stimulated recall review should be viewed as an introspective technique for gaining insight into cognitive processes and implicit beliefs during conceptual change.

Regarding practical aspects, the first implication is that more collaborative argumentation methods should be developed and implemented in science education. In current science learning, science teachers typically use question-answer-evaluation sessions or individual argumentation, rather than taking advantage of the benefits of combining argumentative activities and student-student collaboration (Kilinc et al., 2017; Scott et al., 2006). In addition, science teachers should note the second practical implication of the current findings: that argumentative dialogue should be taken into consideration when collaborative argumentation is designed and implemented. Using collaborative argumentation did not mean that science teachers needed to do nothing. Rather, our results suggested that science teachers should provide students with scaffolding, such as discourse instruction, to create productive argumentation in a collaborative argumentation context (Asterhan & Schwarz, 2016). The third practical implication is for the participants of this study—pre-service teachers, if pre-service science teachers are engaged in collaborative argumentation activities during their teaching education period, they can use these activities in their future science classes and encourage their students to collaboratively participate in argumentative practices (Zembal-Saul 2009). Besides, productive verbal dialogue in collaborative argumentation could be utilized for teacher scaffolding of whole classroom and small group student discourse, especially for the primary education stage when students could be too young to engage in collaborative argumentation.

Moreover, from the broader view, this empirical work was a part, although a tiny part, of the wave of the “practice turn” in the new reform of science education (Furtak & Penuel, 2019). There was a recent conversation centralized in the focus of new reform in science education (Furtak & Penuel, 2019; Osborne, 2019; Parsons, 2019; Hammer & Manz, 2019; Larkin, 2019; Penuel & Furtak, 2019; Southerland & Settlage, 2019). The conversation reflected well not only on a range of perspectives and differences in fundamental assumptions held by science educators within the new reform of science education but also on strong spirits and beliefs on the enterprise of science education from the science education community. For that great conversation, this work might provide the science education community with a piece of material or evidence from the conceptual change perspective. As several science educators (e.g., Penuel & Furtak, 2019; Southerland & Settlage, 2019) addressed and claimed within their essays, that conversation just started: as an indicator of a healthy profession, it will be going on and will also be a common space that more and more newcomers holding their perspectives and evidence join in. We believe that all the perspectives and evidence will be helpful to promote the convergence and even concurrence of the focus in this new reform of science education.

7 Conclusion

The current study investigated the value of collaborative argumentation for conceptual change in science education. The findings confirmed that collaborative argumentation had a delayed but long-lasting effect on conceptual change in science education. Furthermore, the findings indicated that collaborative argumentation could provide opportunities for change in cognitive, ontological, epistemological, and intentional aspects. Finally, when we comprehensively considered the roles of verbal dialogue in collaborative argumentation, a surprising finding emerged, suggesting that long-lasting conceptual change was associated with a U-shaped pattern of argumentative dialogue in collaborative argumentation. This surprising finding was also observed in additional argumentative activity. The current findings shed light on the value of collaborative argumentation for long-lasting conceptual change, deepening current understanding of whether conceptual gains from argumentation activities are contingent on a particular type of verbal dialogue enabled by collaborative argumentation.

8 Limitations and Future Research

This study was conducted in the context of scientific concept learning, so all the findings should be taken into consideration the specificity of the context. This raised the possibility that there may be different and even contradictory findings in another context such as in the context of British parliamentary debate.

First, the sample of students in this study was relatively small, and 95.65% of students were female. When evaluating the results of the current study, it should be noted that including different subject groups, such as a larger number of students or gender-balanced groups of students, may have led to different results. Previous studies have suggested that females are more inclined to exhibit deliberative behaviors during social interactions, whereas males are more inclined to exhibit disputative behaviors during social interactions (Asterhan & Schwarz, 2016). Future studies should be conducted to explore the relationship between gender and argumentative dialogue in collaborative argumentation conditions. If one gender is associated with a particular (or pattern of) argumentative dialogue, this individual characteristic could be used to facilitate inhibiting or enabling the generation of a particular (or pattern of) argumentative dialogue. In addition, scholars further denoted those investigations on scientific concept learning also should take into consideration these factors—social and cultural aspects of thinking and diversities of thinking induced by external social and cultural influences, as scientific knowledge was socially constructed by the groups and environments (Leach & Scott, 2003; Mortimer, 1995; Mortimer & El-Hani, 2014). And with that consideration, future studies should be conducted to explore how students with different individual characteristics induced by social and cultural factors (e.g., social relations, peer status, friendship, and local and cultural norms) may engage in collaborative argumentation, and how these students may use their diversities of thinking or understanding to construct understanding of scientific concepts.

Second, students in this study exhibited better conceptual understanding in the phase of the delayed post-test than in the phase of the post-test. Since the topics for the three argumentation activities were selected by most of the students from a list of topics, students' topic interests/preferences may be one factor that may drive them to search for more information and resources about the topics and return better prepared for their delayed post-test 1 month later. Further consideration should be given to the role of students' topic interests/preferences in conceptual change when conducting argumentation activity design. Moreover, since the appropriateness of driving questions should also be taken into consideration, the driving questions for the second and third argumentation activities will be iteratively developed to better support students' science concept learning in future research.

Third, the current study was conducted in the context of online learning because of the significant disruption caused by the COVID-19 pandemic. The online environment may have relieved the tension that often exists in face-to-face interpersonal interactions, providing a more relaxed atmosphere for students to argue with group members using different types of argumentative dialogue. Further consideration should be given to collaborative argumentation in the context of face-to-face learning, carefully examining the characteristics of argumentative dialogue.

Finally, although this study indicated that long-lasting conceptual change was associated with a U-shaped pattern of argumentative dialogue, it was not able to reveal whether the sequential pattern of argumentative dialogue differed in high-performing and low-performing groups. Thus, further work should be conducted to examine the relationships between the sequential pattern of argumentative dialogue and long-lasting conceptual change.

Appendix 1. Example of the collaborative version of worksheet

You are a scientist in the field of biological evolution with a strong passion for species changes under natural selection. There is now a pending historical case that needs to be solved: during the First Industrial Revolution (1760–1840), the original Typica turned black mysteriously.

Please refer to the following (**Three Steps Ready Go**) for a final explanation of this phenomenon:

Sep 1:. From the "**Information and Data on the Changes of Peppered Moths**" provided by the teacher, please think and give **your own explanation**. Meanwhile, **fill out** the **color chart** of the wings of peppered moths in five periods and submit it (**to fill out in the Star of Questionnaires**).

Sep 2:. This is a very difficult task. You can't do it alone. Next, **you can form a "China Team" with two or three scientists to complete it**;

a. The following three teams also **gave their explanations** for the blackening of the peppered moths according to the **"Information and Data on the Changes of Peppered Moths"**.
b. Please discuss **each explanation** together (including explanation from your own team and from other teams) and explain **why** the explanation under discussion makes sense and **to what extent** your evidence can support this explanation? Also, if you think that an explanation doesn't make sense, explain **why** it doesn't make sense, and **to what extent** your evidence at hand can refute that explanation?
c. Please think again and are there **any** other explanations that you have not considered?

Sep 3:. Nice! Now that you have fully discussed the cause of the color change of the peppered moths. Please write down your (**China Team**) final explanation on the left side of the collaborative writing platform and **submit** the color **change chart** of the wings of the five-generation peppered moth **again in the Star of Questionnaires**.

Appendix 2. Example of the individual version of worksheet

You are a scientist in the field of biological evolution with a strong passion for species changes under natural selection. There is now a pending historical case that needs to be solved: during the First Industrial Revolution (1760–1840), the original Typica turned black mysteriously.

Please refer to the following (**Three Steps Ready Go**) for a final explanation of this phenomenon:

Sep 1:. From the "**Information and Data on the Changes of Peppered Moths**" provided by the teacher, please think and give **your own explanation**. Meanwhile, **fill out** the **color chart** of the wings of peppered moths in five periods and submit it (**to fill out in the Star of Questionnaires**).
Sep 2:. This is a very difficult task. **You need to refer to explanations from other scientists**;

a. The following three scientists also **gave their explanations** for the blackening of the peppered moths according to the **"Information and Data on the Changes of Peppered Moths"**.

Countries of research teams	Explanations for the blackening of peppered moth
	(Australia Team) We believe that the blackening of the peppered moths in Manchester during the First Industrial Revolution is due to **variations induced by soot**: most of the coal used during the industrial revolution was raw coal, with a high degree of coalification and a high carbon content and the emissions tend to be high in carbon, which is in itself an irritant colorant, causing the Typica in the area fully turn black and pass through its genes to the future generations.
	(India Team) We believe that the b Therefore, the peppered moths of each generation after this are black.lackening of the peppered moth in Manchester during the First Industrial Revolution is due to **reproductive options induced by the change of habitats**: the trunk of birch is the habitat of the peppered moths, but when blackened by industrial coal smoke, the Typica is more likely to be found by predators. Therefore, the female Typica seek to mate with the male Carbonaria to make the future generations as black as possible and to avoid the seizure of birds and to finally achieve the continuity of their own gene. The number of Carbonaria increased with the selection of female Typica, while the number of Typica decreased sharply or even died out.
	(Turkey Team) We believe that the blackening of the peppered moth in Manchester during the First Industrial Revolution is due to **the cumulative effects of reproductive differences between two species of peppered moths:** the Typica and the Carbonaria are not two sub-categories of the same species. The Typica are much more reproductive than the Carbonaria. To be concrete, the Typica produce only one generation a year, whereas the Carbonaria produce two to three generations a year. The cumulative effect of reproductive differences is that the number of Carbonaria is increasing, while the scale of Typica is decreasing or even dying out.

b. Please discuss **each explanation** (including explanation from your own and from other scientists') and explain **why** the explanation under discussion makes sense and **to what extent** your evidence can support this explanation? Also, if you think that an explanation doesn't make sense, explain **why** it doesn't make sense, and **to what extent** your evidence at hand can refute that explanation?

c. Please think again and are there **any** other explanations that you have not considered?

Sep 3:. Nice! Now that you have fully discussed the cause of the color change of the peppered moths. Please write down your (**A scientist from China**) final explanation on the left side of the collaborative writing platform and **submit** the color **change chart** of the wings of the five-generation peppered moth **again in the Star of Questionnaires**.

Appendix 3. Examples of conceptual understanding tests

a) *Example from the two-tier item*

You have all heard of the monster Godzilla, but few people know that the prototype of Godzilla is actually an animal called Marine Iguana in nature.

Countries of scientists	**Explanations for the blackening of peppered moth**
	(A scientist from Australia) I believe that the blackening of the peppered moths in Manchester during the First Industrial Revolution is due to **variations induced by soot**: most of the coal used during the industrial revolution was raw coal, with a high degree of coalification and a high carbon content and the emissions tend to be high in carbon, which is in itself an irritant colorant, causing the Typica in the area fully turn black and pass through its genes to the future generations.
	(A scientist from India) I believe that the blackening of the peppered moth in Manchester during the First Industrial Revolution is due to **reproductive options induced by the change of habitats**: the trunk of birch is the habitat of the peppered moths, but when blackened by industrial coal smoke, the Typica is more likely to be found by predators. Therefore, the female Typica seek to mate with the male Carbonaria to make the future generations as black as possible and to avoid the seizure of birds and to finally achieve the continuity of their own gene. The number of Carbonaria increased with the selection of female Typica, while the number of Typica decreased sharply or even died out.
	(A scientist from Turkey) I believe that the blackening of the peppered moths in Manchester during the First Industrial Revolution is due to **the cumulative effects of reproductive differences between two species of peppered moths:** the Typica and the Carbonaria are not two sub-categories of the same species. The Typica are much more reproductive than the Carbonaria. To be concrete, the Typica produce only one generation a year, whereas the Carbonaria produce two to three generations a year. The cumulative effect of reproductive differences is that the number of Carbonaria is increasing, while the scale of Typica is decreasing or even dying out.

Marine Iguana: When Darwin arrived in the Galapagos Islands in South America in 1835, he discovered that there was a marine iguana that was different from land iguana, even though they shared a common ancestor. Unlike land iguana, marine iguanas are formidable swimmers that are able to dive deep into the ocean, hold their breath for a long period, and feed on seaweed.

(1). Individual marine iguana shares some common characteristics, but the degree and size of the common characteristics are different.

Right/Wrong. Explain:____

(2). The physiological changes of marine iguanas are due to the need to swim on their own.

Right/Wrong. Explain: ____

(3). Some marine iguanas have good hearing, and although this trait does not endow them a survival advantage, it is passed on genetically to their future generations.

Right/Wrong. Explain: ____

(4). Marine iguanas have acquired the skill to swim faster by wagging their tails quickly which can be passed on to future generations.

Right/Wrong. Explain: ____

(5). Marine iguanas that are at a disadvantage in the competition for food are more likely to starve to death or suffer from malnutrition at an early age.

Right/Wrong. Explain: ____

b) *Example from the open-ended construction item*

Cheetahs are known as "sprinters".

When a cheetah chases its prey, it can run up to 95 km/h or even faster. But ancestors of cheetah ran quite slowly with the fastest speed of only 30 km/h. How does natural selection lead to this change? Please reveal and describe the process of change in your own words.

____

Appendix 4. Examples of coloring task in the first argumentation task

Please complete the following coloring task of the five-generation peppered moths to depict the gradual change in wing coloring in the five periods of Industrial Revolution (please use a pencil to shade the circles to represent the magnitude of change in peppered moth coloring)

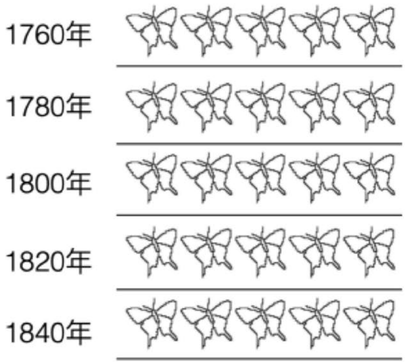
When completing the coloring task, please take photos of your work and uploaded the photos to the Star of Questionnaires platform.

Appendix 5. Examples of background materials for argumentation activities

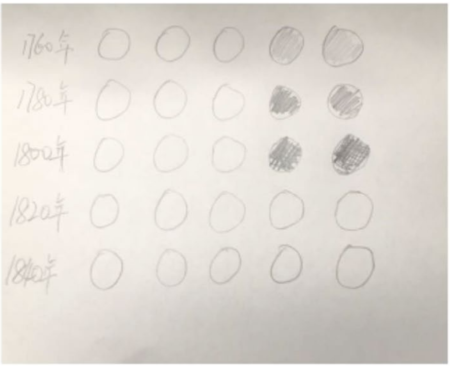
Peppered Moth

With its English name of peppered moth, it can be further divided into two sub-categories: one is called **Typica** with white torso and wings and black spots and the other is called **Carbonaria** with black torso and wings.

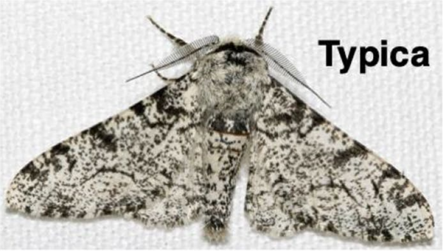
The original version of colouring item

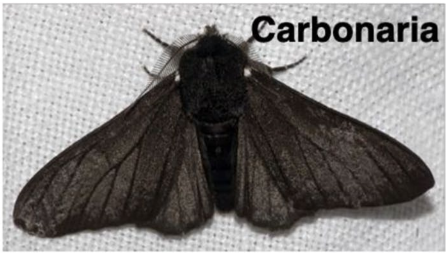


Pencil colouring demonstration



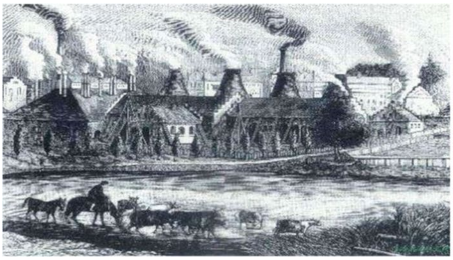
Peppered moths are widespread in temperate countries such as the UK and China. In the UK, they produce only one generation a year. They habituate themselves to flying at night, and perching motionless on the trunks of lichens covered birch trees during the day. Birds are their natural enemies.

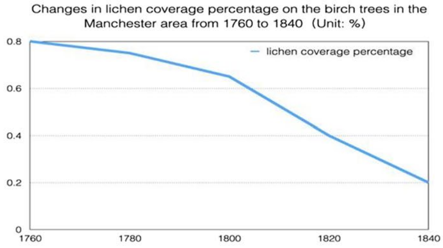


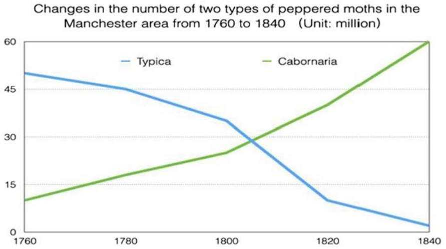


The British Industrial Revolution began in the 1760s, that is, the 1760, when a large number of factories appeared. As "the hometown of the British Industrial Revolution", Manchester has the earliest and largest number of cotton textile factories in the UK. As a great number of coals are needed for the steam engine to weave the cotton, towering chimneys are built in every cotton textile factory, from which thick black soot was discharged directly into the atmosphere without filtration to immerse Manchester in blackness. The birch trees here are also deeply affected. The bark of the birch tree is covered with thick black soot on which the peppered moths are perched.

By 1840, entomologists had discovered that there were more and more Carbonaria in the area, and that there were fewer Typica on the ground. The following are two sets of data: one reflects changes in lichen coverage percentage on the birch trees in the Manchester area from 1760 to 1840 (see the left picture), and the other reflects changes in the number of two types of peppered moths in the Manchester area over the same period (see the right picture).







Acknowledgements This study was supported by the National Natural Science Foundation of China (Grant No: 61877003) and the International Joint Research Project of Faculty of Education of Beijing Normal University. Authors would like to gratefully acknowledge the National Natural Science Foundation of China and the Faculty of Education of Beijing Normal University for their support of this work.

Author Contribution All authors contributed to the study conception and design. The first draft of the manuscript was written by Xiaoshan Li. The first draft of the manuscript was revised by Yanyan Li. Material preparation, data collection, and analysis were performed byWenjing Wang. And all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This paper was supported by the National Natural Science Foundation of China (Grant No: 61877003) and the International Joint Research Project of Faculty of Education of Beijing Normal University.

Data and Material Availability The data and material used in the paper is available from the authors upon request.

Code Availability The code used in the paper is available from the authors upon request.

Declarations

Ethics Approval This study has gained the approval from the ethical committee of the Faculty of Education at Beijing Normal University.

Conflict of Interest The authors declare that they have no conflict of interest.

References

Anderson, C. W., & Smith, E. L. (1987). Teaching science. In *Richardson-Koehler, Educators' Hand- book: A research perspective* (pp. 84–111). Longman, Inc..

Archila, P. A., Molina, J., & Truscott de Mejía, A. M. (2020). Using historical scientific controversies to promote undergraduates' argumentation. *Science & Education, 29*, 647–671.

Asterhan, C. S. (2013). *Epistemic and interpersonal dimensions of peer argumentation* (pp. 251–271). Affective learning together.

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

Asterhan, C. S., & Dotan, A. (2018). Feedback that corrects and contrasts students' erroneous solutions with expert ones improves expository instruction for conceptual change. *Instructional Science, 46*(3), 337–355.

Asterhan, C. S., & Resnick, M. S. (2020). Refutation texts and argumentation for conceptual change: A winning or a redundant combination? *Learning and Instruction, 65*, 101265.

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

Asterhan, C. S., & Schwarz, B. B. (2016). Argumentation for learning: Well-trodden paths and unexplored territories. *Educational Psychologist, 51*(2), 164–187.

Berland, L. K., & Hammer, D. (2012). Framing for scientific argumentation. *Journal of Research in Science Teaching, 49*(1), 68–94.

Breitmayer, B. J. (1991). Triangulation in qualitative research: Issues of conceptual clarity and purpose. *Qualitative nursing research: A contemporary dialogue, 19*(2), 226–239.

Bryman, A. (2016). *Social research methods*. Oxford university press.

Bybee, R. W. (2008). Scientific literacy, environmental issues, and PISA 2006: The 2008 Paul F-Brandwein lecture. *Journal of Science Education and Technology, 17*(6), 566–585.

Carnevale, P. J., & Probst, T. M. (1998). Social values and social conflict in creative problem solving and categorization. *Journal of Personality and Social Psychology, 74*(5), 1300.

Carter, N., Bryant-Lukosius, D., DiCenso, A., Blythe, J., & Neville, A. J. (2014). The use of triangulation in qualitative research. *In Oncology Nursing Forum, 41*(5), 545–547.

Champagne, A. B., Gunstone, R. F., & Klopfer, L. E. (1985). Effecting changes in cognitive structure among physics students. In L. West & L. Pines (Eds.), *Cognitive structure and conceptual change* (pp. 163–187). Orlando: Academic Press.

Chang, C. J., Chang, M. H., Chiu, B. C., Liu, C. C., Chiang, S. H. F., Wen, C. T., ... and Chen, W. (2017). An analysis of student collaborative problem solving activities mediated by collaborative simulations. Computers & Education, 114, 222-235.

Chi, M. T. H. (1992). Conceptual change within and across ontological categories: Examples from learning and discovery in science. In R. Giere (Ed.), *Cognitive models of science: Minnesota Studies in the Philosophy of Science* (pp. 129–186). Minneapolis, MN: University of Minnesota Press.

Chi, M. T. H. (1997). Creativity: Shifting across ontological categories flexibly. In T. B. Ward, S. M. Smith, R. A. Finke, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 209–234). Washington, DC: American Psychological Association.

Chi, M. T. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *The Journal of the Learning Sciences, 14*(2), 161–199.

Chi, M. T. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. *International handbook of research on conceptual change, 61*, 82.

Chi, M. T., Slotta, J. D., & De Leeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction, 4*(1), 27–43.

Corbin, J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Sage.

diSessa, A. A. (1988). Knowledge in pieces. In G. Forman & P. Pufal (Eds.), *Constructivism in the computer age* (pp. 49–70). Lawrence Erlbaum Associates, Inc..

Dreyfus, A., Jungwirth, E., & Eliovitch, R. (1990). Applying the "cognitive conflict" strategy for conceptual change—some implications, difficulties, and problems. *Science Education, 74*(5), 555–569.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84*(3), 287–312.

Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education, 25*(6), 671–688.

Duit, R. H., & Treagust, D. F. (2012). Conceptual change: Still a powerful framework for improving the practice of science instruction. In *In Issues and challenges in science education research* (pp. 43–54). Springer.

Erduran, S. (2007). Methodological foundations in the study of argumentation in science classrooms. In *In Argumentation in science education* (pp. 47–69). Springer.

Evagorou, M., & Osborne, J. (2013). Exploring young students' collaborative argumentation within a socioscientific issue. *Journal of Research in Science Teaching, 50*(2), 209–237.

Felton, M., Garcia-Mila, M., & Gilabert, S. (2009). Deliberation versus dispute: The impact of argumentative discourse goals on learning and reasoning in the science classroom. *Informal Logic, 29*(4), 417–446.

Ferrari, M., & Chi, M. T. (1998). The nature of naive explanations of natural selection. *International Journal of Science Education, 20*(10), 1231–1256.

Fransen, J., Kirschner, P. A., & Erkens, G. (2011). Mediating team effectiveness in the context of collaborative learning: The importance of team and task awareness. *Computers in Human Behavior, 27*(3), 1103–1113.

Furtak, E. M., & Penuel, W. R. (2019). Coming to terms: Addressing the persistence of "hands-on" and other reform terminology in the era of science as practice. *Science Education, 103*(1), 167–186.

Gass, S. M., & Mackey, A. (2013). *Stimulated recall methodology in second language research*. Routledge.

Goldman, A. (1999). *Knowledge in a social world*. Clarendon Press.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64–74.

Hall, B. H., Griliches, Z., & Hausman, J. A. (1984). Patents and R&D: Is there a lag? (No. w1227). *National Bureau of Economic Research*, 1–35.

Hammer, D., & Manz, E. (2019). Odd ideas about learning science: A response to Osborne. *Science Education, 103*(5), 1289–1293.

Heng, L. L., Surif, J., & Seng, C. H. (2015). Malaysian students' scientific argumentation: Do groups perform better than individuals? *International Journal of Science Education, 37*(3), 505–528.

Inagaki, K., & Hatano, G. (2003). *Conceptual and linguistic factors in inductive projection: How do young children recognize commonalities between animals and plants* (pp. 313–334). Language in mind.

Isohätälä, J., Näykki, P., Järvelä, S., & Baker, M. J. (2018). Striking a balance: Socio-emotional processes during argumentation in collaborative learning interaction. *Learning, Culture and Social Interaction, 16*, 1–19.

Jensen, M. S., & Finley, F. N. (1996). Changes in students' understanding of evolution resulting from different curricular and instructional strategies. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 33*(8), 879–900.

Jiménez-Aleixandre, M. P. (1992). Thinking about theories or thinking with theories?: A classroom study with natural selection. *International Journal of Science Education, 14*(1), 51–61.

Johnstone, A. H. (1991). Why is science difficult to learn? Things are seldom what they seem. *Journal of Computer Assisted Learning, 7*(2), 75–83.

Johnstone, A. H. (2000). Teaching of chemistry-logical or psychological? *Chemistry Education Research and Practice, 1*(1), 9–15.

Katchevich, D., Hofstein, A., & Mamlok-Naaman, R. (2013). Argumentation in the chemistry laboratory: Inquiry and confirmatory experiments. *Research in Science Education, 43*(1), 317–345.

Kaya, E. (2013). Argumentation Practices in Classroom: Pre-service teachers' conceptual understanding of chemical equilibrium. *International Journal of Science Education, 35*(7), 1139–1158.

Kilinc, A., Demiral, U., & Kartal, T. (2017). Resistance to dialogic discourse in SSI teaching: The effects of an argumentation-based workshop, teaching practicum, and induction on a preservice science teacher. *Journal of Research in Science Teaching, 54*(6), 764–789.

Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing, 44*, 100450.

Larkin, D. B. (2019). Attending to the public understanding of science education: A response to Furtak and Penuel. *Science Education, 103*(5), 1294–1300.

Leach, J., & Scott, P. (2003). Individual and sociocultural views of learning in science education. *Science & Education, 12*(1), 91–113.

Liu, Q. T., Liu, B. W., & Lin, Y. R. (2019). The influence of prior knowledge and collaborative online learning environment on students' argumentation in descriptive and theoretical scientific concept. *International Journal of Science Education, 41*(2), 165–187.

Luque, M. L. (2003). The role of domain-specific knowledge in intentional conceptual change. In *Intentional conceptual change* (pp. 140-177). Routledge.

McLure, F., Won, M., & Treagust, D. F. (2020). A sustained multidimensional conceptual change intervention in grade 9 and 10 science classes. *International Journal of Science Education, 42*(5), 703–721.

Mercer, N. (1996). The quality of talk in children's collaborative activity in the classroom. *Learning and Instruction, 6*(4), 359–377.

Mercer, N. (2000). *Words and minds: How we use language to think together*. Psychology Press.

Mortimer, E. F. (1995). Conceptual change or conceptual profile change? *Science & Education, 4*(3), 267–285.

Mortimer, E. F., & El-Hani, C. N. (2014). Heterogeneity of thinking and speaking. In R. Gunstone (Ed.), *Encyclopedia of Science Education*. Springer.

National Research Council (NRC). (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.

NGSS Lead States. (2013). Next generation science standards: for states, by states. Retrieved from http://www.nextgenscience.org. Accessed 8 Apr 2021.

Nussbaum, E. M. (2008). Using argumentation vee diagrams (AVDs) for promoting argument-counterargument integration in reflective writing. *Journal of Educational Psychology, 100*(3), 549.

Osborne, J. (2019). Not "hands on" but "minds on": A response to Furtak and Penuel. *Science Education, 103*(5), 1280–1283.

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in science classrooms. *Journal of Research in Science Teaching, 41*(10), 994–1020.

Parsons, E. R. (2019). Why not an integrative and inclusive approach—hands on and "minds on?" A lesson for mentoring 21st century science education researchers. *Science Education, 103*(5), 1284–1288.

Penuel, W. R., & Furtak, E. M. (2019). Science-as-practice and the status of knowledge: A response to Osborne. *Science Education, 103*(5), 1301–1305.

Piaget, J. (1954). *The construction of reality in the child*. Basic Books.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*(2), 211–227.

Reed, C., & Long, D. (1998). Persuasive monologue. In H. V. Hansen, C. W. Tindale, & A. V. Colman (Eds.), *Argumentation and rhetoric* (pp. 1–17). Ontario Society for the Study of Argumentation.

Rich, V. J. (2009). Clinical instructors' and athletic training students' perceptions of teachable moments in an athletic training clinical education setting. *Journal of Athletic Training, 44*(3), 294–303.

Ryu, S., & Sandoval, W. (2008). *Interpersonal influences on collaborative argument during scientific inquiry*. Paper Presented at the American Educational Research Association (AERA).

Sampson, V., & Clark, D. (2009). The impact of collaboration on the outcomes of scientific argumentation. *Science Education, 93*(3), 448–484.

Sampson, V., & Clark, D. B. (2011). A comparison of the collaborative scientific argumentation practices of two high and two low performing groups. *Research in Science Education, 41*(1), 63–97.

Schwarz, B. B. (2009). Argumentation and learning. In N. Muller-Mirza & A.-N. Perret-Clermont (Eds.), *Argumentation and education—Theoretical foundations and practices* (pp. 91–126). Springer Verlag.

Scott, P. H., Mortimer, E. F., & Aguiar, O. G. (2006). The tension between authoritative and dialogic discourse: A fundamental characteristic of meaning making interactions in high school science lessons. *Science Education, 90*(4), 605–631.

Shtulman, A. (2006). Qualitative differences between naïve and scientific theories of evolution. *Cognitive Psychology, 52*(2), 170–194.

Sinatra, G. M., & Pintrich, P. R. (2003). The Role of Intentions in Conceptual Change Learning: Gale M. Sinatra and Paul R. Pintrich. In *Intentional conceptual change* (pp. 10–26). Routledge.

Sinatra, G. M., & Pintrich, P. R. (2003). *Intentional conceptual change*. Routledge.

Sinatra, G. M., & Taasoobshirazi, G. (2011). Intentional conceptual change: The self-regulation of science learning University of Nevada, Las Vegas. In *In Handbook of Self-Regulation of Learning and Performance* (pp. 217–230). Routledge.

Smith, K., Johnson, D. W., & Johnson, R. T. (1981). Can conflict be constructive? Controversy versus concurrence seeking in learning groups. *Journal of Educational Psychology, 73*(5), 651.

Southerland, S. A., & Settlage, J. (2019). An invitation into an ongoing conversation: Revealing different perspectives on a few fundamental assumptions of the work of Science Educators. *Science Education, 103*(5), 1277–1279.

Taber, K. S. (2013). Revisiting the chemistry triplet: drawing upon the nature of chemical knowledge and the psychology of learning to inform chemistry education. *Chemistry Education Research and Practice, 14*(2), 156–168.

Talanquer, V. (2011). Macro, submicro, and symbolic: The many faces of the chemistry "triplet". *International Journal of Science Education, 33*(2), 179–195.

Thagard, P. (1992). Analogy, explanation, and education. *Journal of Research in Science Teaching, 29*(6), 537–544.

van Eemeren, F. H., & Grootendorst, R. (1984). *Speech acts in argumentative discussions: A theoretical model for the analysis of discussions directed towards solving conflicts of opinion* (pp. 12). Dordrecht: Floris Publications

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction, 4*(1), 45–69.

Vosniadou, S. (2003). *Exploring the relationships between conceptual change and intentional learning* (pp. 377–406). *Intentional conceptual change*.

Vosniadou, S., & Mason, L. (2012). Conceptual change induced by instruction: A complex interplay of multiple factors. In *APA educational psychology handbook, Vol 2: Individual differences and cultural and contextual factors.* (pp. 221-246). American Psychological Association.

Vygotsky, L. S. (1978). Socio-cultural theory. *Mind & Society*, 52–58.

Walton, D. (2009). Argumentation theory: A very short introduction. In *In Argumentation in artificial intelligence* (pp. 1–22). Springer.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46*(1), 71–95.

Woods, J. (2004). Appeal to force. In *In The Death of Argument* (pp. 75–94). Springer.

Yang, W. T., Lin, Y. R., She, H. C., & Huang, K. Y. (2015). The effects of prior-knowledge and online learning approaches on students' inquiry and argumentation abilities. *International Journal of Science Education, 37*(10), 1564–1589.

Zembal-Saul, C. (2009). Learning to teach elementary school science as argument. *Science education, 93*(4), 687–719.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Xiaoshan Li[1] · **Yanyan Li**[1] · **Wenjing Wang**[2]

Xiaoshan Li
lixiaoshan@mail.bnu.edu.cn

Wenjing Wang
wenjingwang@mail.bnu.edu.cn

[1] School of Educational Technology/Smart Learning Institute, Beijing Normal University, Xinjiekouwai Street, Haidian District, Beijing 100875, China

[2] School of Systems Science, Beijing Normal University, Xinjiekouwai Street, Haidian District, Beijing 100875, China