Title page

Authors and affiliations

Oliver Scheuer[1]
Saarland University, Saarbrücken, Germany

Bruce M. McLaren
Saarland University, Saarbrücken, Germany *and*
Carnegie Mellon University, Pittsburgh, PA, U.S.A.

Armin Weinberger
Saarland University, Saarbrücken, Germany

Sabine Niebuhr
Clausthal University of Technology, Clausthal, Germany

Manuscript title

Promoting Critical, Elaborative Discussions through a Collaboration Script and Argument Diagrams

Abstract

During the past two decades a variety of approaches to support argumentation learning in computer-based learning environments have been investigated. We present an approach that combines argumentation diagramming and collaboration scripts, two methods successfully used in the past individually. The rationale for combining the methods is to capitalize on their complementary strengths: Argument diagramming has been shown to help students construct, reconstruct, and reflect on arguments. However, while diagrams can serve as valuable resources, or even guides, during conversations, they do not provide explicit support for the discussion itself. Collaboration scripts, on the other hand, can provide direct support for the discussion, e.g., through sentence openers that encourage high quality discussion moves. Yet, students often struggle to comply with the rules of a script, as evidenced by both the misuse and nonuse of sentence openers. To try to benefit from the advantages of both of these instructional techniques, while minimizing their disadvantages, we combined and experimented with them within a single instructional environment. In particular, we designed a collaboration script that guides student dyads through a process of analyzing, interrelating and evaluating opposing positions on a contentious topic with a goal to jointly generate a well-reasoned conclusion. We compare a baseline version of the script, one that only involves argument diagramming, with an enhanced version that employs an additional peer critique script, implemented with sentence openers, in which student pairs were assigned the roles of a proponent and a constructive critic. The enhanced version of the script led to positive effects: student discussions contained a higher number of elaborative moves and students assessed their argumentation learning more positively.

Keywords

computer-supported collaborative learning; collaboration scripts; peer-critique script; argumentation; argument mapping; adaptive support

Citation

Scheuer, O., McLaren, B. M., Weinberger, A., & Niebuhr, S. (2013). Promoting critical, elaborative discussions through a collaboration script and argument maps. *Instructional Science, 41*(3) May 2013 doi: 10.1007/s11251-013-9274-5

[1] Address correspondence to Oliver Scheuer, DFKI, CeLTech, Campus D3.2, D-66123 Saarbrücken, Germany, Phone: +49 681 302 71072, E-mail: o.scheuer@mx.uni-saarland.de

Promoting Critical, Elaborative Discussions through a Collaboration Script and Argument Diagrams

Abstract: During the past two decades a variety of approaches to support argumentation learning in computer-based learning environments have been investigated. We present an approach that combines argumentation diagramming and collaboration scripts, two methods successfully used in the past individually. The rationale for combining the methods is to capitalize on their complementary strengths: Argument diagramming has been shown to help students construct, reconstruct, and reflect on arguments. However, while diagrams can serve as valuable resources, or even guides, during conversations, they do not provide explicit support for the discussion itself. Collaboration scripts, on the other hand, can provide direct support for the discussion, e.g., through sentence openers that encourage high quality discussion moves. Yet, students often struggle to comply with the rules of a script, as evidenced by both the misuse and nonuse of sentence openers. To try to benefit from the advantages of both of these instructional techniques, while minimizing their disadvantages, we combined and experimented with them within a single instructional environment. In particular, we designed a collaboration script that guides student dyads through a process of analyzing, interrelating and evaluating opposing positions on a contentious topic with a goal to jointly generate a well-reasoned conclusion. We compare a baseline version of the script, one that only involves argument diagramming, with an enhanced version that employs an additional peer critique script, implemented with sentence openers, in which student pairs were assigned the roles of a proponent and a constructive critic. The enhanced version of the script led to positive effects: student discussions contained a higher number of elaborative moves and students assessed their argumentation learning more positively.

Keywords: computer-supported collaborative learning; collaboration scripts; peer-critique script; argumentation; argument mapping; adaptive support

It is widely recognized that critical thinking skills play an important role in today's information societies. Being able to understand, evaluate, and produce arguments in a well-reasoned way is crucial in many professions (e.g., science, the law, politics), in contributing to democratic society, and in our private lives. During the past two decades many computer-based tools have been developed to support the acquisition of argumentation skills (e.g., Kirschner et al. 2003; Scheuer et al. 2010; Noroozi et al. 2012; Scheuer et al. 2012).

Following Bell (1997) two classes of systems can be distinguished: *Argument representation systems* support students in collecting, classifying and organizing argumentation knowledge. Tools such as Reason!Able (van Gelder 2002), Belvedere (Suthers 2003), and LARGO (Pinkwart et al. 2009) support students in analyzing philosophical, scientific and legal arguments, respectively, through diagrammatic representations. By using these systems students learn about domain-specific argument structures and criteria to assess the quality of arguments. *Discussion-based systems*, on the other hand, support students in conducting argumentative dialogues. These systems are based upon special-purpose communication interfaces designed to promote pedagogically valuable forms of discourse between students (Soller 2001; McAlister et al. 2004; McLaren et al. 2010; Stegmann et al. 2012). By using these systems students learn how to use arguments skillfully in discussions to convince others of a position or to collaboratively explore a topic or problem. For instance, the Argunaut system (McLaren et al. 2010) supports students in collaboratively discussing and arguing thorny ethical issues, such as whether it is ethical to perform experiments on animals.

In this paper, we present an instructional approach that connects argument representation with discussion activities to promote critical, elaborative dialogues. The approach utilizes a hybrid software system that supports both argument diagramming and scripted argumentative discourse, capitalizing on their complementary strengths: making argument structures visible and promoting productive forms of peer discussion. The collaboration script, in which students take "proponent" or "critic" roles during their discussions, has been implemented in the tool's communication interface through *sentence openers*, that is, predefined phrases students can choose from when composing chat messages that are sent between them. The sentence openers have been designed precisely to promote the "proponent" and "critic" roles. For instance, a proponent may provide an illustrative example to support one of his points by starting a message with the sentence opener "For instance,..." A critic may raise an objection against one of the partner's statements by starting a message with the sentence opener "But..." In a study we investigated the central research question: "Will a proponent / critic role script, used in conjunction with argument diagrams, lead to higher quality discussions and more learning?"

To motivate our research – and formulate our research questions in a precise manner – we start with a review of research on argument diagramming and scripted discourse. We next present preliminary work in the form of a first version of our approach and promising results from a study using this initial version. We then discuss in detail an enhanced version of the instructional approach, the software system used to support the approach, and an associated empirical study. We conclude the paper with an analysis of aspects of the system and the pedagogical approach that could be improved through *adaptive support*, for instance, by having the instructional system encourage critical comments from students when they have focused too much on the positive aspects of a specific perspective.

Argument Diagramming

Argument diagramming (also called *argument mapping*) is an approach to graphically representing the structure of arguments. Typically, arguments are decomposed into their constituent propositions (e.g., claims, evidence) and relations between propositions (e.g., a piece of evidence supports or opposes a claim), and laid out in the form of node-and-link graphs. The labels attached to nodes and links may differ depending on the specific argumentation domain. Besides field-independent schemes, such as the Toulmin model (1958), specialized notational systems (or ontologies) have been proposed to represent, for instance, legal arguments (Wigmore 1931), or arguments about planning and design problems (Rittel and Webber 1973). Since the mid-1990s a number of computer-based argumentation systems based on the argument diagramming paradigm have been developed for educational purposes, for instance, Belvedere (Suthers et al. 1995), Reason!Able (van Gelder

2002), and Digalo (Schwarz and Glassner 2007). Such tools have been used in a variety of ways and in different domains, for instance, to analyze existing legal arguments (Pinkwart et al. 2009), to outline arguments in preparation for essay writing (Erkens et al. 2005; Janssen et al. 2010), and to discuss a given contentious question (Schwarz and Glassner 2007; McLaren et al. 2010).

The main feature of argument diagrams is that argument structures are represented visually and explicitly. In contrast to less explicit formats, such as prose, diagrams allow students to immediately see how lines of reasoning evolve, step by step, without having to infer argumentative relations (van Gelder 2005). By way of their explicitness diagrams can help students *see* faulty reasoning. While prose and chat are linearly arranged, argument diagrams allow multilevel hierarchies (or networks), thus better match the hierarchical structure of many arguments (van Gelder 2005). Cognitive processing can be further facilitated through graphical elements such as colors, lines and shapes (van Gelder 2002). While expressing knowledge in a highly structured format unavoidably involves cognitive overhead, e.g., through a "premature commitment to structure," it has the potential to trigger processes of reflection and deeper understanding (Buckingham Shum et al. 1997). In particular, the specific notations and category systems used in argument diagramming tools can focus students' attention on important concepts of argumentation and encourage reflection about these concepts, e.g., the distinction between hypotheses and data in scientific arguments (Suthers 2003).

On the downside, argument diagrams may become unwieldy, especially in synchronous collaborative settings when many contributions are created in rapid succession (Scheuer et al. 2010). Sometimes the result is a "spaghetti" image, which is hard to read and follow (Loui et al. 1997). In general, the quality and readability of argument diagrams depends on how skillfully users organize and spatially arrange information. Diagrams arranged according to Gestalt principles, such as symmetry, continuation and proximity, have been shown to be more useful learning resources, leading to higher learning gains (Dansereau 2005). Another factor is the complexity of the notational language used. Students may struggle choosing between many categories that only differ in subtle ways. Notational languages should therefore avoid any unnecessary complexity and focus on differentiations important with respect to the learning goals to minimize distraction and extraneous cognitive load. Suthers et al. (2001), for instance, significantly simplified the initial notational language used in Belvedere, since students had difficulties in grasping the differences in meaning between some of the categories (e.g., whether a hypothesis "predicts" or "explains" a piece of data). Finally, while the typical arrangement of argument diagrams according to logical and thematic relationships helps students to focus on the underlying argument structure, the temporal sequence of contributions is less clear compared to chat or threaded discussions, making it harder to identify recent contributions, a problem that may become critical when diagrams are used as a discussion medium, a not uncommon approach in Computer-Supported Collaborative Learning settings (Scheuer et al. 2010). Notably, many of these issues can be remediated or softened through system functions, for instance, orientation support (e.g., mini-maps, search functions), and awareness support (e.g., displaying creation timestamps, highlighting recent contributions).

In terms of learning, positive effects of the argument diagramming method have been shown in a variety of domains. Twardy (2004) and Harrell (2008) report significantly higher learning gains in terms of critical thinking skills (Twardy) and argument analysis skills (Harrell) for introductory philosophy courses in which argument diagramming was taught. Pinkwart et al. (2009) report on two evaluation studies of LARGO, an intelligent tutoring system for legal argumentation, in which students map legal argument transcripts into argument diagrams. In a first study, conducted with volunteer first-semester law students, low-aptitude students especially benefited from using LARGO compared to students using a text-based annotation tool. However, the results were not replicated in a follow-up study, in which students were required to use LARGO as a compulsory part of the course. Easterday et al. (2009) tested how diagrams affect learning in the domain of policy deliberation. In an immediate posttest they found that providing causal diagrams along with text significantly improves students' ability to make correct inferences about that text. In a transfer test, in which students had to make correct inferences from a given text without diagrams or diagramming software being available, those who used the diagramming software during a learning phase outperformed those who did not use the diagramming software. Janssen et al. (2010) compared argument diagrams and a list-based format as representational tools for jointly analyzing a historical debate. Diagram users constructed better representations of the debate, wrote essays of higher quality, and learned more (i.e., pre-to-posttest knowledge gains). Suthers and Hundhausen (2003) investigated the impact of three different knowledge representation formats (Graph, Matrix, Text) on collaborative learning in the context of discussions about science and public health problems. The results provide evidence in line with the theory of representational guidance (Suthers 2003), which postulates a connection between the saliences and constraints imposed by specific representational formats on the one hand and learning on the other hand. In particular, students in the Graph and Matrix conditions elaborated more on previously discussed ideas compared to the Text condition. Students in the Matrix condition elaborated more on evidential relations compared to both the Text and Graph condition. On closer inspection it turned out that many of the evidential relations students considered in the Matrix condition were irrelevant. It appears that the strong prompting character of the matrix – the matrix invites one to think about a possible relation for each cell, that is, for each and every hypothesis/data pair – might have tempted students to include many weak relations. In sum, graphs seem to be an appropriate knowledge representation format since they can focus students better on important ideas and relations than text while avoiding the critical shortcomings of matrices, which are generally less expressive than graphs (Scheuer et al. 2010) and possibly induce students to include irrelevant information.

In summary, theoretical arguments and empirical evidence suggest that argument diagramming is an appropriate method to teach argumentation and critical thinking skills. However, to teach how to use sound and convincing arguments in discussions it might be a more suitable approach to engage students directly in discussion, since then they would practice precisely what it is they ought to learn. While argument diagrams can stimulate and guide discussion activities (Suthers 2003), they provide less explicit – and possibly less effective – support for productive discussion processes as compared to scripted discourse approaches, discussed next.

Scripted Discourse

A major focus of the field of Computer-Supported Collaborative Learning (CSCL) has been dialogic forms of argumentation. A number of approaches have been investigated to support student discussions by way of structured communication interfaces that implement *micro-scripts*, that is, dialogue models that describe desirable discussion moves and sequences (Dillenbourg and Hong 2008). Similar to notational languages employed in argument diagrams, these communication interfaces provide a form of representational guidance. They raise students' awareness of productive discussion moves and encourage their use (Baker and Lund 1997). The basic rationale is that students incrementally internalize behaviors embodied in the micro-script and transfer these behaviors to situations in which no support is available.

There are different approaches to micro-scripts. One is to use argument diagrams for structured discussions rather than knowledge representation (McLaren et al. 2010; Dragon et al. in press). In graphical e-Discussions, students post new messages by creating labeled boxes in a shared workspace (e.g., an argument, a question, a clarification), and reply to existing messages by connecting boxes through labeled arrows that denote different kinds of argumentative or rhetorical relations (e.g., for, against, relates to). Schwarz and Glassner (2007) investigated the effects of structuring graphical e-Discussions through floor control (i.e., a graphical user interface that guides turn-taking) and an informal ontology (i.e., node and link labels derived from informal conversations). The structuring led to discussions of higher quality with more relevant claims and arguments, and fewer chat-like contributions (e.g., use of profanity). Stegmann et al. (2007) used an approach that integrates with current Computer Mediated Communication (CMC) practices, since it builds upon a standard online discussion board. Their script comprised two components. The first component aimed at supporting the construction of single arguments. When writing a new message, students filled in a template based on the Toulmin (1958) argumentation model with boxes provided to enter a claim, grounds, and a qualification. The second component aimed at supporting the construction of argumentation sequences through preset message subject lines. The first message in a thread was always labeled "Argumentation," the second one "Counterargumentation," the third one "Integration," the next one again "Counterargumentation" and so on. The script led to process improvements in terms of a better formal quality of argumentation (e.g., more grounded and qualified claims; Weinberger and Fischer [2006]). Also, positive effects on the acquisition of argumentative knowledge have been found. Yet, students' acquisition of domain knowledge was not better when using the script. A study by Jeong and Joung (2007) hints at the possible limitations of explicit message labels. Students instructed to label their messages in an asynchronous discussion board according to prescribed categories, such as argument, evidence, critique, or explanation, challenged fellow students less often than students who were not so instructed. A possible explanation is that participants avoided "critique" messages because they did not want to be overly confrontational with their fellow students (Jeong and Joung 2007).

Another approach, which we have adopted for our research, is the use of *sentence openers* (or note starters), that is, predefined phrases students choose from to start new messages (McManus and Aiken 1995). Typically, students complete these messages in their own words, but in some cases students also choose from a limited set of propositions to complete the message (Baker and Lund 1997). The composition of messages can be further supported through guiding questions (Oh and Jonassen 2007). For instance, to scaffold the creation of a rebuttal, a guiding question might be "What are conflicting issues?" An early example of the sentence opener approach can be found in the Computer-Supported Intentional Learning Environment (CSILE), which dates back to 1983 (Scardamalia and Bereiter 2006). CSILE provides a shared space for representing, organizing and advancing ideas as networks of interlinked notes, comparable to an argument graph. As scaffolds for thinking and discourse, students can use pre-formulated phrases when filling text into the notes, such as "My theory..." and "This theory cannot explain..." Other systems enhance chat and threaded discussion interfaces with sentence openers, e.g., Group Leader Tutor (McManus and Aiken 1995), C-CHENE (Baker and Lund 1997), BetterBlether (Robertson et al. 1998), AcademicTalk (McAlister et al. 2004), InterLoc (Ravenscroft 2007) and the Future Learning Environment (FLE3; Oh and Jonassen 2007). Depending on the specific type of dialogue that researchers were hoping to foster, different sets of sentence openers have been used. InterLoc (Ravenscroft 2007), for instance, can support multiple types of dialogue through corresponding sentence openers, among others, critical discussions and creative reasoning dialogues. The dialogues are modeled as *dialogue games* and formalized in terms of participant roles, dialogue moves, corresponding sentence openers, and rules of interactions. Since sentence openers can be mapped to specific communicative intentions, computers can achieve some level of dialogue understanding without complex natural language processing and understanding (Baker and Lund 1997). InterLoc (Ravenscroft 2007) capitalizes on this by recommending sentence openers appropriate to respond to previous contributions, based on the rules of the underlying dialogue game. The Group Leader Tutor (McManus and Aiken 1995) uses sentence openers to support and diagnose specific collaboration skills (e.g., leadership, creative conflict) in the context of collaborative problem solving, based on a model that associates sentence openers with these collaboration skills (Johnson and Johnson 1991).

While sentence openers have the potential to shape student interactions in favorable ways, there are limitations and challenges that also must be considered. On the one hand, sentence openers can reduce students' typing load since frequently used text fragments can be added with a click of a button (Baker and Lund 1997; Lazonder et al. 2003; Soller 2001). On the other hand, sentence openers must be carefully and systematically organized in the user interface, possibly grouped according to higher-level categories, to help students quickly find appropriate sentence openers (Baker and Lund 1997; Soller 2001). The set of available sentence openers should be broad enough to satisfy students' communicative needs and avoid misuse. Soller (2001), for instance, found that only 68% of all messages started with sentence openers that matched the intentions expressed in the message body. In response, she extended the set of available sentence openers to offer more choices. On the other hand, too many options reduce the salience of individual sentence openers and increase search time. As pointed out by Lazonder et al. (2003) in reference to a study by McManus and Aiken (1996), students may avoid the difficulty of classifying the content of their messages by always picking an unspecific sentence opener such as "I think..." One way to reduce the use of overly general or inappropriate sentence openers is to allow free-text messages in addition to scripted ones (Baker and Lund 1997). Yet, this has the potential to undercut the goals of the script as well when students consistently ignore the provided sentence openers and use the interface like a standard chat. Lazonder et al. (2003) found that students who can choose between sentence openers and free-text strongly prefer the free-text option (only 8% of on-task messages used a sentence opener), a preference that increases with the amount of chat experience.

To date, relatively little empirical evidence has been published regarding the pedagogical effectiveness of sentence opener approaches. In a small-scale investigation, Baker and Lund (1997) found that students using a sentence opener interface produced more task-focused and reflective messages. Nussbaum et al. (2002) report that students who used a sentence opener interface disagreed significantly more with one another compared to students who used the standard version of a bulletin board. The effect varied in strength depending on certain personality traits such as curiosity (openness to ideas) and anxiety (measured through items such as "I often worry about things that might go wrong"). Sentence openers stimulated particularly those with lower levels of curiosity to critically engage with other opinions, while anxious students did not benefit as much from the provided structuring. McAlister et al. (2004) found that discussions structured through sentence openers were more focused (i.e., less off-topic contributions) and contained a greater number of reasoned claims and critical responses (rebuttals and utterances of direct disagreements). A possible explanation is that students interpreted the presence of sentence openers as "permission" to critically respond (McAlister et al. 2004). Interestingly, these results run counter to the ones reported by Jeong and Joung (2007) who found that messages labeled as "critique" inhibited rather than promoted the use of critical responses. Oh and Jonassen (2007) report somewhat mixed results. Students using the sentence openers interface provided significantly more evidence than ones who used the standard threaded discussion interface. Also, the number of messages that contributed to the problem-solving process was significantly higher on several coding dimensions (e.g., hypothesis testing messages). On the other hand, students in the threaded discussion condition used significantly more verification messages, which are about determining the value of arguments, an important activity as well. In sum, the presented evidence clearly shows that sentence openers can bias discussion behaviors to the better (task focus, reflectiveness, critical engagement, evidence and reasons). The contradictory results presented by Jeong and Joung (2007) suggest that the specific conditions of use may play a role (e.g., student characteristics, instructional framing and context). Overall, the somewhat limited amount of reported data calls for further investigation of the sentence opener approach.

Some of the decisions in the design of sentence opener interfaces mentioned above, such as the breadth of options, the availability of a free-text option, and automated, system-generated feedback, point to a more general tradeoff widely discussed in the CSCL literature, the tradeoff between *freedom* and *guidance* in instructional scripts (Dillenbourg 2002). On the one hand, scripts can provide structures that guide students in accomplishing specific tasks by encouraging (or enforcing) productive behaviors and discouraging (or eliminating) unproductive / counterproductive behaviors. On the other hand, the provided structures can be a limiting factor and prevent productive behaviors not anticipated in the script. This problem is particularly relevant when it comes to open tasks that can be addressed in many, unforeseeable and personalized ways. Collaboration and argumentation processes certainly fall into this category. CSCL micro-scripts may guide argumentation processes based on a specific model of desired interaction. Yet, this model may conflict with a student's' personal approach, which might be appropriate as well, leading to frustration on part of the student and thus hindering task progress and learning. For instance, the richness of natural languages offers manifold ways to express one and the same idea. The stylized language of sentence openers may conflict with a student's personal preferences in expressing ideas. Or there is simply no adequate sentence opener available that suits a specific communicative intention (Soller 2001). Then again, a loosely defined script, one that grants high degrees of freedom, may invite students to disregard and not behave according to the script, e.g., to avoid additional efforts, so the instructional goals of the script are not achieved. For instance, Lazonder et al. (2003) could not evaluate the effectiveness of their sentence opener approach because students strongly preferred creating and sending free-text messages, i.e. sentence openers were only rarely used. In sum, it is important to find the right tradeoff between the two poles of coerciveness and freedom. Unfortunately, we cannot expect a "one-size-fits-all" solution, that is, a script with the "right" level of guidance for *all* students in *all* situations. Novice learners might benefit from more guidance, while more experienced learners might benefit from more freedom. As students become more able and independent problem-solvers (or arguers) one may want to reduce support and transfer responsibilities to the learner, an approach referred to as *fading the scaffold* (Pea 2004). Technical realizations of automated script adaptations like fading the scaffold are still a major research challenge since they require a sufficiently precise diagnosis of student, group or situational parameters to regulate the adaptation of the script (Scheuer et al. 2012).

External representations can play a pivotal role in collaborative learning arrangements. For instance, within the C-CHENE system, student dyads create and discuss diagrammatic models of storage, transfer and transformation of energy (Baker and Lund 1997). This leads to intriguing reasons to *combine* argument diagrams with discussion activities. Argument diagrams can facilitate communication (Buckingham Shum et al. 1997), function as an intelligible group memory (Buckingham Shum et al. 1997), and serve as resources, stimuli and guides in conversations (Suthers 2003). A number of questions regarding the use of argument diagrams in collaborative learning settings have been addressed, for instance, the effects of creating diagrams individually before a debate versus collaboratively during a debate (Munneke et al. 2003), the effects of using diagrams as discussion medium, as opposed to using them to represent a preceding discussion (Lund et al. 2007), and ways to integrate diagrammatic and typewritten discourse (Suthers et al. 2008). In our research, we investigate whether enhancing a chat interface with sentence openers can improve student discussions about a controversial topic that has been represented in an argument diagram previously. The sentence openers are part of a peer critique script that has been designed to stimulate critical and elaborate discussions.

Research Questions

In light of the prior research just discussed, our rationale for combining *structured representations* (i.e., argument diagrams) and *structured discourse* (i.e., sentence openers) is as follows. As discussed above, argument diagramming is a valuable learning activity in and of itself and has been shown to be effective in improving argumentative reasoning skills. However, our primary goal is to support students in engaging in *argumentative discussions*, that is, the skilled use of arguments to jointly and critically elaborate and evaluate different positions. As learned from prior research (e.g., Buckingham Shum et al. 1997; Suthers 2003), diagrams can facilitate discussions in several respects. They provide students with background knowledge in the form of an "inventory" of relevant arguments to enable fruitful discussions. They visually represent the most important statements and argumentative relations, thus, focusing students on the "argumentative gist" of the texts. Yet, students may struggle to make appropriate and effective use of the knowledge represented in the diagram during their conversations. The structure provided by sentence openers may help them to better capitalize on the representational affordances of diagrams by promoting the conversion from a graphical into a conversational knowledge representation.

Vice versa, the use of structured communication interfaces like sentence openers may be facilitated through the structured representation of knowledge in argument diagrams. Past research has yielded mixed results regarding the effectiveness of scripted discourse approaches. Some results suggest that scripts can improve the quality of discussions (Nussbaum et al. 2002; McAlister et al. 2004; Stegmann et al. 2007), while other results suggest that students may have problems accommodating the structures dictated by the script, leading to misuse (Soller 2001) and non-use (Lazonder et al. 2003). Pre-structuring of knowledge in the form of an argument diagram (and the mental structures emerging while the diagram is created) may encourage and help students to use sentence openers in a more informed and appropriate way. For instance, a statement labeled as an "argument" in the diagram and linked through a "supports" arrow to some other statement, may readily translate into the use of a sentence opener like "An argument for this point is..." A piece of knowledge labeled as "data" in the diagram may prompt a student to use a related sentence opener like "According to a statistic / estimate…" In sum, we hope for synergistic effects of combining diagrams and scripted discourse.

The study reported in this paper gives a partial answer to this question by keeping the "diagram" factor constant, while varying the "scripted discourse" factor. Our baseline is a well-structured sequence of activities that involves argument diagramming and discussion. We investigate whether the *addition* of a peer critique script, implemented through corresponding sentence openers and instructions, can improve process and outcomes as compared to the baseline design. More specifically, we address the following questions:

- RQ1: To what extent do students make use of the given sentence openers?
- RQ2: Does the peer critique script improve the quality of online discussions in terms of the depth of elaboration?
- RQ3: Do students assess their learning more positively when using the peer critique script?
- RQ4: Does the peer critique script improve students' factual knowledge of the discussed topic?

Preliminary Work

We devised two versions of a collaboration script we have called **FACT** ("**F**ostering **A**rgumentation through **C**onflicting **T**exts") to support critical, elaborative forms of argumentation. In this section we briefly discuss the design, rationale and results of FACT-1, an initial, testable implementation of our approach and preparatory work for the study discussed in this paper. FACT-2, the main focus of the current article, will be discussed in the next section. While FACT-1 does not use either argument diagrams[2] or sentence openers, it allowed us to test how important parameters the main study is based upon work out in reality. As discussed below, key elements of FACT-1 have been brought forward to FACT-2, yet implemented in a more advanced way.

FACT-1 Script

Andriessen and Schwarz (2009) proposed research on how to support group discussions with argumentative texts as an important future direction in CSCL. Moreover, they proposed that great potential lies in the use of texts that support conflicting positions. The FACT-1 script is based on such an approach, utilizing opposing texts to stimulate student discussions. FACT-1 provides three key elements:

E1: Prompting individual preparation. Past research has shown that successful collaboration usually involves a combination of individual and collaborative activities (Baker 2003; Jermann and Dillenbourg 2003; Rummel and

[2] We initially planned to integrate argument diagram activities, but ultimately could not do so because of technical problems.

Spada 2005). Individual preparation gives students time to make up their own minds about a controversial issue without social pressure (Baker 2003). It also allows students to develop their own ideas before the ideas of others influence their thinking. Thus, more diverse knowledge resources can be activated and contributed to collaborative argumentation (Weinberger et al. 2007). With a clear picture on a given topic in mind, gained from careful individual deliberation, students are better prepared to engage in fruitful interaction with others.

E2: Creating conflict. Proponents of the socio-cognitive conflict theory see the attempt to resolve social disagreements as a key component of cognitive development and conceptual learning (Doise and Mugny 1984). To create conflict, we let students first make a decision between two alternatives in the individual preparation phase; in the collaborative decision phase, we pair up students with opposite opinions. To emphasize initial disagreement, we make students aware of their different decisions. Conflicting opinions call for explanations, justifications and collaborative conflict resolution – activities that have been shown to be supportive of learning (Nussbaum 2008). Similar tactics aimed at inducing and emphasizing conflict to promote discussions and learning have been used, for instance, in teaching physics (Baker 2003; Clark et al. 2009) and instructional design (Jermann and Dillenbourg 2003).

E3: Encouraging productive collaboration and discussion norms. The final element is a set of guiding instructions students are asked to read before collaborating with one another. We try to promote the following "productive" behaviors: (1) a mutual commitment to the starting point (Van Eemeren and Grootendorst 2004). Students should be aware of what their partners know, believe and argue for. Thus, we ask students to peer-review and discuss the results of the individual phase. (2) Willingness to criticize the position of others. "Consensual" groups often achieve suboptimal results compared to "critical" groups (Postmes et al. 2001). We encourage students to take a critical position by identifying and discussing possible weaknesses in their partner's contributions. (3) Constructive synthesis. We attempt to scaffold collaborative writing by encouraging the following: agreeing on the main thesis, agreeing on the main points supportive of the thesis, agreeing on the distribution of work, and finally, in iterative cycles, writing, peer-reviewing and discussing answer components.

Note that *E2* (creating conflict) and *E3* (encouraging productive collaboration) only appear, at first consideration, to be at odds with one another. Conflict is part of productive collaboration, in which concerns and objections must be identified and addressed in a rational way.

Empirical Results

To investigate the effects of the FACT-1 script on student collaboration, an empirical study was conducted with undergraduate students at Carnegie Mellon University (U.S.) (Scheuer et al. 2011). The study pursued the hypothesis that the three structuring elements discussed above would lead to increased elaboration of content (e.g., more arguments, counterarguments and explanations). The study used a quasi-experimental design with 38 participants (19 dyads) distributed over two conditions (Script condition: 11 dyads; Control condition: 8 dyads). Control dyads collaborated freely while collaboration in Script dyads was structured through the FACT-1 script. In both conditions, students collaborated on different, networked computers.

To stimulate lively and critical discussions climate change ethics was used as the discussion topic. Climate change is a complex and controversial problem involving uncertainty and ethical ramifications. As with many other real-world problems there are multiple well-reasoned perspectives and no formally correct solution. The specific thesis to be discussed was "Developed countries have to cut their carbon emissions drastically." Two argumentative texts with conflicting conclusions were selected. The *Lomborg* text argues for only moderate emission reductions based on a cost-benefit analysis. Poor countries could be helped more effectively if money is spent otherwise. The *Brown* text argues for substantial emission reductions based on the ethical obligations by first-world countries.

The study intervention spanned two sessions. Pairs of students read the two conflicting texts before the first session took place. During the sessions Google Documents (http://docs.google.com), with its collaborative document editor and integrated chat tool, was used as the collaboration platform. In the first session students answered three essay questions, in which they summarized important aspects of the two texts and chose the position they found more compelling. While Control dyads answered these questions collaboratively, Script dyads worked individually (cf. *E1*). Before the second session took place Script students, who had not collaborated until this point, were grouped together to form dyads. In order to promote conflict, each dyad was composed of one student favoring Lomborg's position and the other student favoring Brown's position (cf. *E2*). In this session students worked on another essay question that asked them to agree on and reasonably justify a joint position. Control dyads continued, as before, in free collaboration. Script dyads received additional guiding instructions according to *E3*.

The results generally confirmed the hypothesis – the script was successful. In particular, a significantly higher number of quality discussion moves (argumentative broadening and deepening) were found in the chats between the members of the Script dyads (Scheuer et al. 2011). From the literature we know that conflict schemes, that is, composing student groups in a way that maximizes opinion conflicts, can have positive effects on the quality of student discussions (Jermann and Dillenbourg 2003; Clark et al. 2009). Similarly, structuring collaboration through scripted activity sequences has been shown to be conducive to problem solving and learning (e.g., Rummel and Spada 2005). Since we tested the script as a whole, we cannot definitely say whether one specific script element, or the combination of multiple script elements, made the difference. Based on the literature we hypothesize that each element made some contribution to the result. There is also reason to believe that the different elements positively interacted with one another. Typically, the different elements comprising a collaboration script are designed and combined to *collectively* support specific interaction patterns. In this specific instance, the individual preparation gave students time to develop a clear and well-reasoned opinion through a deliberate process. That is, students could gain more confidence in their opinions, and consequently, feel better prepared and more comfortable to argue in favor of that position in a dispute with a learning partner. So the individual preparation may have amplified the effect of the conflict scheme.

On the other hand, the overall quality of discussions was relatively low in both conditions. Students often focused more on the actual problem solving (i.e., answering of the given essay question and coordination of the writing process) rather than *jointly and deeply* elaborating on subject matters, an observation also made by Baker

(2003) in a similar setup, in which students also had to simultaneously take on a discussion and collaborative writing task.

The FACT-2 Script

Based on the results of the preliminary study (as well as further pilot tests) the FACT-2 collaboration script was developed. FACT-2 combines argument analysis and argumentative discourse activities. Students first individually create diagrams to analyze the argumentative structure of given texts and then collaboratively discuss these texts based on the created diagrams. Below we describe the LASAD diagramming environment and the FACT-2 collaboration script.

LASAD Argument Diagramming Environment

LASAD (Loll et al. 2012) is a web-based argument-diagramming environment; only a standard web browser is required to use the system. Student groups create argument diagrams in a shared workspace to represent the structure of arguments (see Fig. 1, panel on the left). Boxes represent statements and links represent argumentative and rhetorical relations between statements (e.g., "support", "opposition", "related to"). In contrast to many other argument diagramming systems, LASAD is highly configurable. For instance, labels and visual appearance of boxes and links can be configured. To associate additional information with boxes and links, GUI widgets such as text fields, radio buttons or dropdown menus can be added. Additional tools and displays can be added (e.g., a chat tool, a panel showing the list of active users, a panel displaying texts to be analyzed). An analysis service can be configured to automatically detect relevant patterns in argument diagrams (e.g., cyclic arguments, incorrectly used links, keywords in text fields) and to provide feedback (highlighting of patterns and textual messages) (Scheuer and McLaren 2013).

Fig. 1 shows the specific LASAD setup used in the instructional approach of our study. Students choose between four box types ("main thesis", "main argument", "helping argument" and "fact") and two link types ("support" and "opposition"). This relatively simple and informal setup allows students to quickly grasp how to use the tool. Some prior research has shown that complex representational schemes can lead to confusion on part of the students and are therefore often more detrimental than beneficial in guiding students' thinking and interactions (Suthers et al. 2001). Moreover, judgments regarding the importance of statements are explicitly and visually represented ("main arguments" versus "helping arguments"), so important statements can be identified at a glance. The panel on the right shows a sentence opener interface, one of the tools that can be added to a LASAD setup. The buttons in the lower right allow students to choose from a predefined set of sentence openers to start the next chat message; the highlighted openings in the actual chat show previously used sentence openers. These phrases scaffold and encourage the use of messages considered conducive to fruitful discussions (e.g., to ask questions or to be critical). Sentence openers are part of the experimental treatment and described in more detail in section "Study." (A comparison group used a standard text chat instead.)

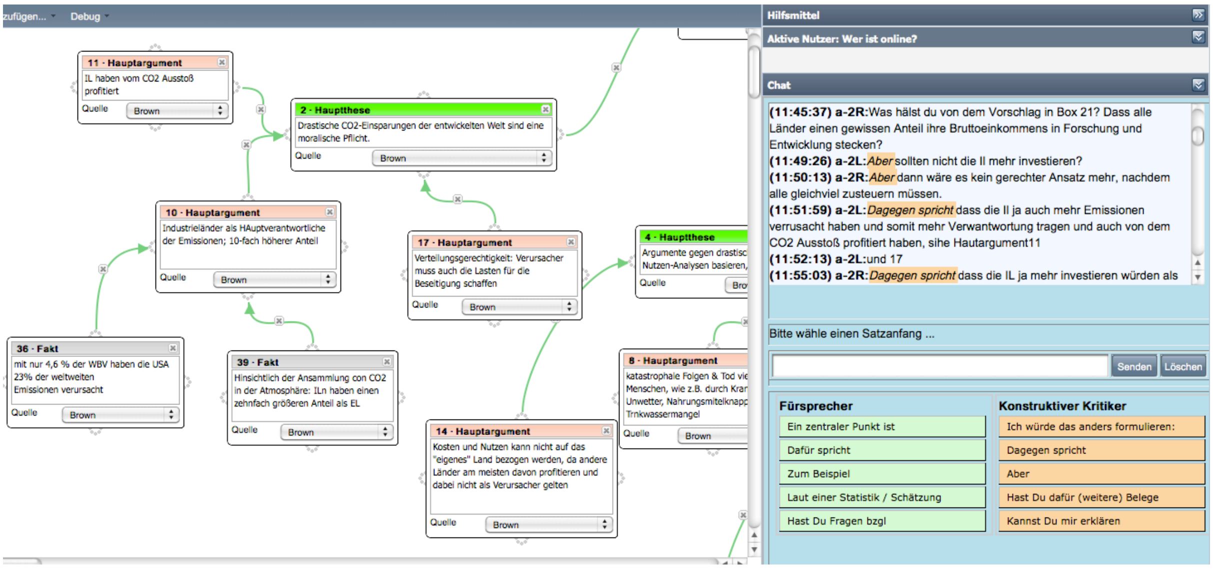

Fig. 1: LASAD argument diagramming system (setup used in the experimental treatment)

FACT-2 Collaboration Script

FACT-2 advances the concept of FACT-1 in several respects. First, an argument-diagramming environment is used, with the advantages discussed above (Buckingham Shum et al. 1997; Suthers 2003; Scheuer et al. 2010). Second, a distributed resources design is used, that is, each student has exclusive access to one of the two texts. When resources are distributed, each student depends on the contributions of the partner to achieve the best possible solution ("positive resource interdependence"), a condition known to be beneficial for collaboration and learning when students are striving for a common goal (Johnson and Johnson 2009). Finally, phases and activities have been redefined as depicted in Table 1. In phase 1, students individually read and analyze the given texts and represent the respective lines of argumentation in a diagram. In phase 2, students discuss, based on the diagrams, aspects of the individual texts with their partner. In phase 3, students discuss relationships between the two texts. In phase 4, students agree on a joint position (which can be one of the positions in the texts or some compromise between the positions) and compose a joint, reasoned conclusion. The script allows some degrees of flexibility, i.e., students can use some time in a new phase to conclude the activities of the previous phase. Overall, the phases represent a progression from structured, narrowly focused activities to more free form, open activities.

Table 1: Phases of the FACT-2 collaboration script

Phase	Description
1. Analyzing (individual)	– read assigned text – model argument in a LASAD diagram
2. Discussing (collaborative)	– model argument in a LASAD diagram (cont'd) – discuss individual texts
3. Interrelating (collaborative)	– discuss individual texts (cont'd) – discuss relations (e.g., conflicts, agreements) between texts
4. Concluding (collaborative)	– agree on a joint position – write down a justified joint conclusion

The three script elements described in the context of FACT-1 (*E1–E3* in section "Preliminary Work") are also addressed in FACT-2, yet, in a somewhat different form:

- E1: Individual preparation takes place in phase 1, in which students individually analyze the given texts.
- E3: Productive collaboration and discussion norms are supported through (a) the sequence of phases (and accompanying instructions), which guide students to systematically analyze, interrelate, and evaluate the given texts to ultimately synthesize a single, shared position, (b) the argument diagrams, which create a shared focus on the epistemic structure of the given texts (e.g., assumptions, claims, and arguments) (Suthers 2003), thus, encouraging epistemic rather than relational forms of conflict resolution ("being critical of ideas, not people" [Johnson and Johnson 1994]), and (c) the distributed resources design, which emphasizes group interdependence and cooperation over individualized activities and competition (Johnson and Johnson 2009).
- E2: Conflict is created through (a) conflicting texts (Andriessen and Schwarz 2009) and (b) a peer-critique script that promotes the roles of a proponent and a constructive critic (Weinberger et al. 2005), which was investigated in that study and will now be described.

Study

The FACT-1 script required a 50/50 split among students between both positions to form dyads with conflicting opinions. Since this is not always possible we developed a peer-critique script for FACT-2 in which students are assigned roles, a proponent of their text and a constructive critic of their partner's text, independent of their own opinion. A sentence opener interface (Baker and Lund 1997; Soller 2001) is used to provide support for the proponent and constructive critic roles (see Table 2 and Fig. 1).

Our general research hypothesis is that the peer-critique script leads to more elaborative and critical discussions, which in turn leads to better learning. To test this hypothesis we conducted a quasi-experimental study at Saarland University, Germany on July 8th – 9th, 2011, using a pretest-intervention-posttest design.

Table 2: Sentence openers used in the chat tool version of the experimental treatment (translated from German)

Proponent	**Constructive Critic**
A central point is …	I would phrase this differently: …
An argument for this point is …	An argument against this point is …
For instance, …	But …
According to a statistic / estimate …	Do you have any / more evidence …
Do you have questions regarding …	Could you explain to me …

Sample

Participants were students enrolled in the Humanities and Social Sciences at a German university. Participation was voluntary and advertised in several ways (e.g., announcement in Philosophy classes, on Philosophy and

Psychology mailing lists). Students received a participation fee of 35 Euros and could earn another 5 Euros for each additional student they convinced to participate (i.e., brokerage fee).

Overall 46 students registered for one of four offered sessions. Two sessions comprised the *Script+* condition (i.e., script *with* peer-critique component; 12 dyads); the other two comprised the *Control* condition (i.e., script *without* peer-critique component; 10 dyads). The majority of participants were female (68%). A more detailed characterization per condition is provided below in section "Homogeneity of Condition."

Procedure

Fig. 2 depicts the sequence of activities. The study procedure started with a pretest and a training task to familiarize students with the LASAD diagramming system. After a 10-minute break, students worked on the actual task as described in Table 1. Before the first collaborative phase (P II), the Script+ condition received short additional instructions regarding the peer-critique script (3 minutes). The actual task was followed by a 5-minute break before a posttest was administered.

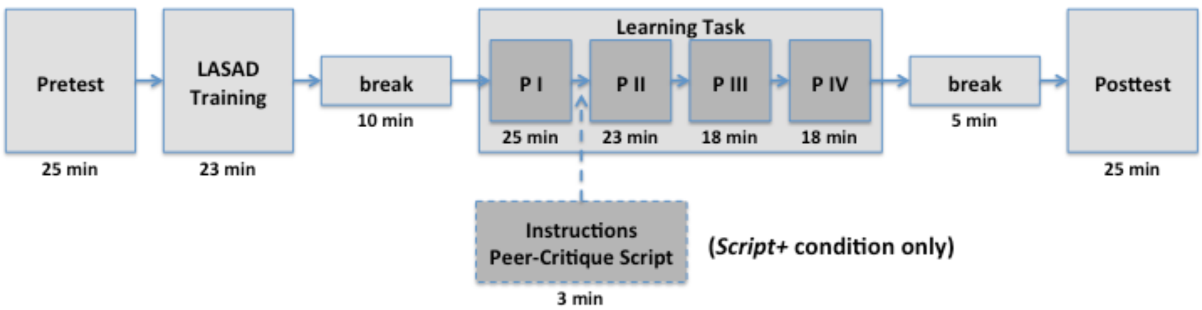

Fig. 2: Sequence of activities during study

Materials and Instruments

The learning task was based on the same topical background (climate ethics) and materials (Brown and Lomborg) as the preliminary study described above. Yet, rather than using the original texts, students received a 3-page summary of their text at the beginning of the learning task. Moreover, based on other Brown writings, the Brown summary was enriched with a discussion of several shortcomings of cost-benefit arguments in the global warming discussion in order to better align the texts. Instructions regarding the script phases and peer-critique script were provided orally and on paper.

Table 3: Pretest items: Preferences regarding Learning Environment and Task (translated from German)

Variable	*Description*	*Example Item*
interest-in-climate-change	Interest in the given knowledge domain (climate change)	"I'm interested in the topic of climate change."
pref-collab-learning	Preference for collaborative learning (as compared to individual learning)	"I prefer learning on my own."

pref-text-chat	Preference for computer-based chat tools	"I don't like internet chats much."
pref-visual-presentations	Preference for visually represented information (as compared to textually represented information)	"I prefer diagrams and charts over detailed oral or written explanations."
pref-argumentation	Preference for argumentative problem resolution	"I like to discuss things extensively."

Web-based questionnaires were administered before and after the learning task. The pre-questionnaire elicited socio-demographic data (*age*, *gender*, *number-of-semesters*). A 5-point Likert-scale was used to measure learning environment and task variables that plausibly have an influence on students' behavior and performance: *interest-in-climate-change*, *pref-collab-learning*, *pref-text-chat*, *pref-visual-representations* and *pref-argumentation* (3 items per variable). The Likert scales used the following levels (translated from German): "Don't agree (-2)," "Somewhat don't agree (-1)," "Neutral (0)," "Somewhat agree (1)," and "Agree (2)." Data elicited in the pre-questionnaire was used to test for homogeneity of conditions. Examples are provided in Table 3.

The post-questionnaire variables measured how students assessed their own learning during the session: *perceived-learning-climate-ethics*, *perceived-learning-arg-theory* (i.e., "knowing" about argumentation) and *perceived-learning-arg-practice* (i.e., "doing" argumentation) using the same 5-point Likert-scale. Examples of post-questionnaire self-assessment items are provided in Table 4.

Table 4: Posttest items: Self-assessment of Learning (translated from German)

Variable	*Description*	*Example Item*
perceived-learning-climate-ethics	Perceived learning of contents in the given knowledge domain (climate ethics)	"In this online session I learned something about the topic of climate ethics."
perceived-learning-arg-theory	Perceived learning of argumentation generally ("knowing about" argumentation)	"In this online session I learned something about argumentation."
perceived-learning-arg-practice	Perceived learning of argumentation practices ("doing" argumentation)	"In this online session I learned to argue better."

Nine multiple-choice questions (MCQs) were administered as part of the post-test to assess students' knowledge of the two texts. Five MCQs focused on the Brown text and four on the Lomborg text. For each MCQ students had to select between four options to complete given Lomborg / Brown statements. Examples of Brown and Lomborg knowledge items are provided in Table 5.

Table 5: Posttest item: Knowledge of Texts (translated from German)

Variable	*Description*	*Example Item*
knowledge-of-lomborg-statements	Detailed factual knowledge about contents of the Lomborg text	"Lomborg: The *precautionary principle*, if understood as a legal principle, says that: **(a)** we have an obligation to take all possible precautionary measures against climate change **(b)** a lack of full scientific certainty about climate change alone is

		no excuse to take no precautionary measures **[correct answer]** **(c)** all countries should take equal measures against climate change to achieve a maximum of safety **(d)** the ones who caused climate change have to pay for caused damages since they failed to take precautionary measures in advance"
knowledge-of-brown-statements	Detailed factual knowledge about contents of the Brown text	"Brown: The Intergovernmental Panel on Climate Change (IPCC): **(a)** estimated that the emissions from 1990 must be drastically reduced to stabilize the amount of greenhouse gases in the atmosphere at a 'safe' level **[correct answer]** **(b)** withheld concrete numbers regarding a 'safe' global emission level because developing countries feared that CO2 reductions might threat their economic growth **(c)** did not publish concrete numbers regarding a 'safe' global emission level on instigation of the U.S.A. **(d)** in general does not publish concrete numbers regarding a 'safe' global emission level because, as a political body, the IPCC avoids getting involved in scientific controversies"

Analysis Approach

Satisfactory internal consistency values (Cronbach's $\alpha > .70$) based on three items each were achieved for the pre-questionnaire variables *interest-in-climate-change* ($\alpha = .79$), *pref-collab-learning* ($\alpha = .92$), *pref-text-chat* ($\alpha = .93$) and *pref-visual-representations* ($\alpha = .80$). The criterion was met for *pref-argumentation* after dropping one of the three items ($\alpha = .78$). For each variable item scores were aggregated to yield an overall score.

The analysis of students' knowledge gains is based on the nine MCQs in the posttest. Since the elicited information was very specific to the given texts (i.e., specific Lomborg and Brown statements, see Table 5), it is highly unlikely that students possessed any relevant pre-knowledge to answer these questions. Therefore, the analysis does not consider any pre-test knowledge scores. For each student three scores were computed: a Lomborg score, a Brown score, and a total score. Scores were normalized on a scale from 0 to 100.

Chat protocols were coded to assess the amount of elaboration. To assign a proportionally higher weight to more extensive contributions messages were segmented into sentence-level units based on punctuation marks (".", ";", "!", "?"). Using the Rainbow framework (Baker et al. 2007) as a guide, a coding handbook was developed, distinguishing three levels of elaboration (Low, Medium, High). The *Low* category comprises contributions that do not refer to specific subject matter (e.g., off-topic talk, social relation, interaction management, topic-unrelated task management). The *Medium* category comprises contributions that are conducive to knowledge elaboration and relate to specific subject matters, yet without fully elaborating on the contents (e.g., asking a clarification question, proposing a specific topic to discuss, unelaborated opinions and concessions). The *High* category comprises contributions that cite, elaborate, question or criticize relevant contents. Table 6 shows how the coding scheme relates to the original Rainbow framework and gives examples for each category.

The category *High* comprises a relatively broad and heterogeneous set of contributions but is nevertheless instrumental in differentiating between better and worse chats (more High moves means more content was covered in breadth or depth). An advantage of a more coarse-grained coding approach like this is a simpler and more reliable coding procedure. A satisfactory inter-rater agreement (Cohen's $\varkappa > .70$) was achieved between two independent coders ($\varkappa = .76$). For each chat protocol the number of codes per category was aggregated into an overall score ("code-and-count").

Two-sided Student's t-tests were used to compare metric variables between conditions (unpaired t-tests) and between one another (paired t-tests). This also applies to variables based on Likert-scale items, which were considered as interval-scaled. Intra-subject comparisons of scores achieved for Brown and Lomborg MCQs are based on paired t-tests as well. Effect sizes were computed based on Cohen's d. Pearson's Chi-squared test was used to compare the gender distributions of both conditions. All tests used a significance level of .05.

Table 6: Coding scheme, its relation to the Rainbow framework, example messages (translated from German)

Code	*Subsumed Rainbow codes*	*Example chat messages*
Low	1. Outside activity	– "Just as a side note, how do you like this method?"
	2. Social Relation	– "Hello" – "How are you?" – "Thank you."
	3. Interaction Management	– "are you still writing?" – "I cannot find box #57" – "box #106, right hand side"
	4. Task Management [topic-unspecific]	– "which main argument should we discuss first?" – "let's start summarizing our conclusion" – "we still need a justification"
Medium	4. Task Management [topic-specific]	– "let's start with Lomborg's main thesis" – "regarding helping argument #44:"
	5. Opinion	– "what do you think [regarding our last point]" – "could you give me an example for #15?" – "whatever, I think Lomborg is right" – "I agree" – "then I misunderstood something"
High	6. Argumentation	– "emission cuts are not only ethically ineffective but also factually unenforceable" – "Lomborg's main thesis is well supported by #34 and #22" – "Brown is right that ethical factors are not considered, #73"
	7. Broaden & Deepen	– "Brown developed the main ethical idea well, Lomborg found a solution for it in #38" – "well, actually, we should spend money on both issues [climate protection and support of poor countries]"

Results

Before we present the quantitative results to answer the four research questions formulated above, we present and discuss a concrete individual case in detail to give the reader a sense of part of the learning task and how students tackled it.

Case Analysis

Table 7 presents an excerpt from a Script+ chat. We use bold-faced text to indicate the use of sentence openers. In the original user interface, sentence openers are highlighted green (for proponent contributions) and red (for critic contributions). Student S1 read the Brown text (pro drastic emission cuts), student S2 the Lomborg text (anti drastic emission cuts). We selected a case that demonstrates one extreme – a discussion segment in which sentence openers have been consequently used. The excerpt illustrates how our peer-critique script can shape student discussions, including positive aspects and problems. In section "Outlook" we will return to this example to discuss how some problems could potentially be mitigated through *adaptive, system-generated support*.

Table 7: Excerpt from a Script+ condition chat (translated from German)

#	St.	Contribution
1	S1	**A central point is** that climate change can only be controlled through drastic emission cuts
2		**An argument for this point is** that according to an estimate of the IPCC global emissions must be cut by 60% to keep the amount of greenhouse gases in the atmosphere at a safe level
3	S2	**I'd phrase this differently:** Investments into climate protection can cause similar harm.
4	S1	**Do you have (any / more) evidence**
5	S2	**A central point is** that we might forget about other important social projects
6	S1	**A central point is** that industrial countries must accept the main responsibility to solve the climate change problem and to reduce their use of the atmosphere
7	S2	**For instance** there won't be any money for schools, hospitals, etc.
8	S1	**An argument for this point is** that industrial countries are responsible for the main share of emissions
9		**For instance** the USA is responsible for 23 percent
10		**A central point is** that cost-benefit analyses are based on ethically dubious arguments
11	S2	**Do you have (any / more) evidence**
12	S1	**An argument for this point is** that they are biased (e.g., oil industry)
13		**A central point is** that the quantification of costs and benefits of climate protection measures is a general problem

Despite the option of composing free-text messages, the dyad makes *consistent use* of the provided structuring, using a sentence opener in every message without exception. Also, they make *appropriate use* sentence openers, perhaps with the exception of turn #3, in which the sentence opener indicates the modification of a previous statement while the message body is a straight counterargument against a prior statement. In many cases we see evidence of good argumentation behavior. For instance, student 1 substantiates his claim in turn #1 through the data provided in turn #2. Student 2 justifies his claim in turn #3 through the reason given in turn #5, and illustrates this reason through the examples given in turn #7. Students also adopt an informed and academic style of arguing by citing numbers and statistics (turn #2 and #9). Possibly, this behavior was triggered through the explicit representation of "facts" in the argument diagram. Overall, in line with the quantitative analysis below, the discussion is very task-focused without any off-topic talk. Many arguments from the texts are cited and connected to one another.

On the downside, the students' discussion has more breadth than depth, i.e., points are mentioned, briefly explained, but not *critically* elaborated. In turn #3 student 2 objects to his partner's conclusion. From then on, both students are solely focused on supporting their own position without referring to what their partner has said (except for some generic clarification questions, turn #4 and #11). A closer look at the selected sentence openers reveals that, apart from turn #3, only proponent moves are used, which is in accordance with our observation that critical references to the partner's contributions are missing. Also, the arguments stem from copying from the argument analysis with little co-elaboration of *new* meanings and ideas. On the other hand, this clearly shows that the argument diagram has been used as a resource for structured discussions. In sum, while promoting a well-reasoned dialogue, the script was not very successful in engaging students in critical interactions with one another. We will address the question whether this tendency is representative of the whole sample when discussing *RQ1* below.

Homogeneity of Conditions

On average, students in the Script+ (S+) and Control (C) conditions were roughly the same age (S+: 22.5; C: 23; p = .62) and had the same number-of-semesters (S+: 5; C: 5; p = 1). Regarding gender, there were proportionally more female students in the Script+ condition, yet not at a significant level (S+: 75%; C: 45%; p = .16). Differences regarding relevant interests and preferences were relatively small and also not significant: *interest-in-climate-change* (S+: 0.81; C: 0.67; p = .62), *pref-collab-learning* (S+: -0.04; C: 0.08; p = .74), *pref-text-chat* (S+: -0.01; C: 0.32; p = .38), *pref-visual-representations* (S+: 0.08; C: -0.10; p = .54) and *pref-argumentation* (S+: 1.04; C: 1.30; p = .26). It can be concluded that Script+ and Control conditions did not differ significantly in terms of relevant interests and preferences.

RQ1: To what extent do students make use of the given sentence openers?

On average, Script+ students used a sentence opener (SO) in chat messages in one out of five messages (20%), that is, SOs were used considerably more as in the Lazonder et al. (2003) study. The use of SOs differed notably between dyads: five dyads made frequent use (> 25% SO messages), three dyads made occasional use (> 10% SO messages), and four dyads made rare use of SOs (< 10% SO messages). Note that 25% SO messages is already quite considerable since a substantial portion of each dialogue is about greeting, interaction and task management, off-topic talk, etc., rather than about subject matters. In the next section we will see that about one third of all messages in the Script+ chats fall in the category of Low elaboration, which subsumes all discussion moves that are not directly concerned with content-level elaborations. Moreover, even if some dyads did not make heavy use of SOs, the mere presence of SOs in the user interface might have clarified the expectations regarding the two roles to fulfill. That is, students might have freely formulated messages in the spirit of the given SOs. We deliberately decided in favor of such a less coercive scripting approach and gave students the option of sending "standard" chat messages without SOs. For instance, we did not want to force students to use SOs (and thus misuse them) when none of the available options would fit their communicative intentions. In general, finding the "right" level of freedom / coercion is one of the main challenges in the design of CSCL scripts (Dillenbourg 2002).

Critic SOs (Table 2 right column; 54%) were used slightly more often than proponent SOs (Table 2 left column; 46%). This result indicates that the observation we made in the "Case Analysis" – that students rarely used critic SOs – is not representative of the whole sample, and speaks against the hypothesis that students tend to avoid statements that are explicitly marked as criticism (Jeong and Joung 2007). In sum, in light of the quantitative and qualitative results, there is good reason to believe that the script promoted, rather than inhibited, critical interaction – yet not in all dyads to the same extent. We also noticed considerable differences in the use of specific SOs: Three SOs were frequently used (each representing more than 15% of all SO uses): "A central point is…," "But…," and "For instance..." Two SOs were occasionally used (each representing between 10% and 15% of all SO uses): "Could you explain to me…" and "A supporting argument is…" The other five SOs were rarely used (each representing less than 10% of all SO uses; for reference, see Table 2). Interestingly, the SO to elicit whether the partner could follow explanations ("Do you have questions regarding…") was almost never used (about 1% of all SO uses).

With regard to the observation by Lazonder et al. (2003) – that chat experience negatively correlates with a preference for using SOs – we came to a different result. There was a small positive non-significant correlation between the preference for computer-based chat tools and the relative proportion of messages in which a SO was used ($r = .21$; $t = 1.02$; $p = .32$). An analysis of intra-class correlations (ICC) revealed that students within the same dyads significantly influenced each other in terms of the extent of SO uses ($ICC = 0.83$; $F = 10.6$; $p < .001$). That is, if one student decided to use SOs, it was very likely that the other student decided to as well. Thus, if there was an effect of chat preference on the use of SOs, it was essentially superseded by intra-dyad influences and mutual attunement.

RQ2: Does the peer critique script improve the quality of online discussions in terms of the depth of elaboration?

Table 8 shows the results of the chat analysis. Script+ chats were on average shorter, with less Low and Medium elaboration messages and more High elaboration messages, both in terms of absolute and relative numbers. The most interesting result is a non-significant trend ($p = .07$) with large effect size ($d = 0.82$) in terms of High codes in favor of the Script+ condition, a result in accordance with our hypothesis. The difference in terms of the percentage of High codes is highly significant, i.e., students in the Script+ condition were proportionally more engaged in behaviors that we intended to promote. However, this percentage comparison should be taken with a grain of salt since it penalizes chats with many Low or Medium contributions, which are not *per se* indicative of bad collaborative behavior.

Table 8: Statistical summary of chat analysis (Script+ versus Control)

	Script+		**Control**		**Differences**			
	M	**SD**	**M**	**SD**	**abs**	**t**	**p**	**d**
# Total	83.2	29.8	94.8	36.6	-11.6	0.82	.42	-0.35
# Low	31.1	17.3	50.6	28.6	-19.5	1.96	.06	-0.82
# Medium	13.0	4.6	18.6	11	-5.6	1.61	.12	-0.66
# High	38.9	17.9	25.6	14.4	13.3	1.90	.07	0.82
% Low	36	11	54	23	-18	2.36	.03	-0.98
% Medium	16	5	18	10	-2	0.68	.50	-0.29
% High	48	11	28	16	20	3.52	.00	1.48

RQ3: Do students assess their learning more positively when using the peer critique script?

Our hypothesis is further supported by the post-questionnaire analysis, which indicates that Script+ students assessed their learning more positively than Control students. In terms of *perceived-learning-arg-theory* we found a non-significant trend (p = .07), in terms of *perceived-learning-arg-practice* a significant difference (p = .01) favoring the Script+ condition. In terms of *perceived-learning-climate-ethics* we found no difference (p = .26). Across both conditions, students believed they learned most about climate ethics (M = 0.93, SD = 0.97), second most about argumentation (M = 0.23, SD = 1.22), and least about actually participating in argumentation (M = -0.2, SD = 1.11). All differences are highly significant ($p < .001$).

RQ4: Does the peer critique script improve students' factual knowledge of the discussed topic?

On average, students answered slightly more than half of all MCQ questions correctly (M = 54; SD = 20). There was no difference in means in terms of Brown and Lomborg questions (M = 54, SD = 28 [Brown], SD = 32 [Lomborg]).

Table 9 shows a statistical comparison between the Brown and Lomborg item scores for students who were assigned the Brown text (Brown students, first row) and students who were assigned the Lomborg text (Lomborg students, second row). Not surprisingly, students performed significantly better on items that elicited information from "their" text (in both cases: $p < .001$). For MCQs that tested students on the text they modeled in LASAD, they scored, on average, 72 points (for Brown students; SD = 20) and 70 points (for Lomborg students; SD = 31). For the text they only knew from a partner's diagram and discussion with that partner, students scored, on average, 38 points (Brown students on Lomborg questions; SD = 24) and 36 points (Lomborg students on Brown questions; SD = 24). The cross-text scores are well above the chance rate of 25 points but clearly lower than the score obtained for "their" own text. One reason why students did not do better might be that very specific and detailed information was elicited in the questions (e.g., whether an approach proposed by Lomborg would save more or about the same amount of human lives compared to measures in accordance with the Kyoto protocol, see also Table 5). Students might remember such information from a careful reading of the text, but it is less likely they would have discussed that level of detail in their chat discussions. The hypothesis that the discussions contributed little to the achieved MCQ scores is further supported by the fact that the number of High elaboration moves did not correlate with the average MCQ score per dyad (r = .08; t = 0.36; p = .73). Research on knowledge maps (a variation of concept maps [Novak 1990]) suggests that the reading of diagrammatic representations has positive effects on the memory of main ideas rather than on the memory of subordinate ideas (O'Donnell et al. 2002). This could explain why exposure to a partner's diagram did not have a major influence on the MCQ posttest scores for students, since the MCQ test questions were focused more on subordinate facts than the main ideas of the texts.

Table 9: Statistical summary of posttest MCQs analysis (Brown items versus Lomborg items)

	Brown items		**Lomborg items**		**Differences**			
	M	**SD**	**M**	**SD**	**abs**	**t**	**p**	**d**
Brown st.	72	20	38	24	34	5.98	.00	1.55
Lomborg st.	36	24	70	31	-34	5.50	.00	-1.23

Table 10 shows a statistical comparison between Script+ and Control students with respect to MCQ scores. In total, and on both sub-scales, there were no differences between the Script+ and the Control condition. This is in accordance with our assumption above: Students' knowledge of text details is mainly based on the modeling task rather than the discussion task, so the treatment, which only took effect during the discussions, apparently had marginal, if any, influence on this aspect of students' knowledge gains.

Table 10: Statistical summary of posttest MCQs analysis (Script+ versus Control)

	Script+		Control		Differences			
	M	**SD**	**M**	**SD**	**abs**	**t**	**p**	**d**
Total	53	22	55	19	-2	0.28	.78	-0.09
Brown items	53	30	55	27	-2	0.19	.85	-0.06
Lomb items	53	32	55	32	-2	0.19	.85	-0.06

Discussion

Table 11 summarizes the results we obtained to answer our four research questions. Overall, we found evidence in favor of our hypothesis both in the chat protocols and in the participants' assessment of their learning: a peer-critique script, used in combination with an argument diagram, can promote the quality of student discussions. We observed benefits on different dimensions: a trend with large effect size regarding the number of elaborative moves (i.e., arguments, counterarguments and explanations) (Table 11, RQ2), a trend regarding students' assessment whether they learned about argumentation (Table 11, RQ3), and a significant effect regarding students' assessment about whether they learned to argue better (Table 11, RQ3). A closer look at the Script+ condition showed that students indeed utilized the sentence openers to their advantage: two thirds of Script+ dyads made frequent or occasional use of sentence openers (Table 11, RQ1); others may have oriented themselves by reading the given sentence openers while using their own formulations. We observed a substantially increased use of sentence openers compared to the Lazonder et al. (2003) study. A possible explanation is that the diagrams prompted and guided students in selecting appropriate sentence openers. Anecdotal evidence for this has been found in the case analysis: Students reused text fragments from their argument analysis and combined them with appropriate sentence openers to compose chat messages. Thus, combining argument diagrams and sentence openers might indeed have led to benefits that could not have been achieved with either method alone.

No differences were found in terms of detailed knowledge of the given texts (Table 11, RQ4). As expected, students had much better knowledge of the text they read and then modeled themselves as compared to the text they learned about from their partner's modeling and subsequent discussion between the partners. Yet, the approach of reading and modeling one text, while reviewing an opposing text through discussion with a partner may have been beneficial. Rather than focusing on subtle details of one text in isolation, which we tested for in our posttest, the setup of this experiment appeared to prompt students to interrelate and critically evaluate positions and arguments, which may have led to a broader, multi-perspective understanding of the topic at hand and the emergence of new perspectives (Wegerif et al. 2010). Previous results show that engagement in dialectical argumentation indeed promotes deeper processing and conceptual change (Asterhan and Schwarz 2009). Our analysis does not allow for a definitive conclusion on this, which is in general hard to assess due to a wide variety of sometimes surprising and unpredictable insights students gain from fruitful learning discussions.

Future research might address this issue more explicitly, for instance, based on an analysis of an essay-writing task conducted before and after the intervention (e.g., "Make a statement as discerning and well reasoned as possible concerning the question '...' under inclusion of both, supporting and opposing arguments."). Such an analysis could reveal whether students (1) have changed their perspective on the given topic based on evidence and reasoned argument, (2) can provide more substantial support for their perspective, (3) are more willing and capable to acknowledge arguments and facts favoring positions that oppose their own, (4) are aware of different value systems underlying different perspectives, and (5) know about arguments that they did not know before or even ones that were not included in the background readings. A similar question could target a new topic in order to test whether a general attitudinal change towards other perspectives occurred, that is, whether students are more able to see the pros and cons of the different positions, independent of their own initial positions, and, based on this, conceive an informed opinion, which is potentially synthesized from elements of opposing standpoints. This would indicate a step towards an "evaluative epistemological stance," an attitude that is positively linked to argumentative skill development (cf. Kuhn 1991, pp. 172-203). Another question that calls for further investigation is whether students maintain improved argumentation practices when no structural support is available (cf. Pea 2004). This could be tested in a post-intervention application phase (similar to the experimental setup used by Rummel and Spada [2005]), in which students engage in collaborative argumentation using a discussion environment without sentence openers. The resulting chat traces could be analyzed using the same coding approach that was used to code the intervention chats. In addition, it could be checked whether Script+ students make increased use of the sentence openers they had available to them during the intervention.

Table 11: Summary of results regarding the four research questions

Research Question	**Result**
RQ1: To what extent do students make use of the given sentence openers?	– two-thirds of all Script+ dyads made frequent or occasional use of sentence openers – "critic" sentence openers were used slightly more often than "proponent" sentence openers (54% vs. 46%)
RQ2: Does the peer critique script improve the quality of online discussions in terms of the depth of elaboration?	– (non-significant) trend with large effect size in terms of High elaboration codes (in favor of the Script+ condition) – (difference is significant when comparing proportions of High elaboration codes rather than absolute numbers)
RQ3: Do students assess their learning more positively when using the peer critique script?	– significant effect in terms of perceived learning of argumentation practices (in favor of the Script+ condition) – (non-significant) trend in terms of perceived learning of argumentation generally (in favor of the Script+ condition) – no difference in terms of perceived learning about the topic of climate ethics
RQ4: Does the peer critique script improve students' factual knowledge of the discussed topic?	– no difference in terms of detailed knowledge of the given texts

In sum, our results are in line with previous research underscoring that the quality of collaboration / argumentation practices can be improved through appropriate structural support (e.g., McAlister et al. 2004; Rummel and Spada 2005; Weinberger et al. 2010). In particular, peer-critique scripts (e.g., Weinberger et al. 2005; Andriessen et al. 2003; Johnson and Johnson 1994) may help students to engage in positive collaborative behaviors that would otherwise not (or only to a lesser extent) occur.

There is good reason to believe that the FACT-2 script and the design principles it is based upon can also be successfully applied in domains other than climate ethics. As long as there are two (or more) opposing positions that can be represented in a diagram, FACT-2 can be employed to guide the analysis and discussion of these positions. For instance, one could imagine using FACT-2 to incite discussions about other socio-ethical dilemmas, which naturally have different reasonable perspectives, such as the use of genetically modified food (Munneke et al. 2003), experiments on animals (McLaren et al. 2010), or the right to die (Cavalier and Weber 2002). Similarly, FACT-2 can be applied for planning and design problems, which often require the reconciliation of opposing design concerns. For instance, engineering students could investigate the tradeoffs between technical efficiency, financial costs, and environmental friendliness in the design of technical systems (Chaudhuri et al. 2009). Instructional science students could review and discuss the controversy on the usefulness of collaboration scripts as an instructional instrument as opposed to more free-style forms of collaboration (Dillenbourg 2002). Other possible application areas are manifold, including politics, history, science, medical decision-making, and the Law. Depending on the specific domain and pedagogical objectives, certain adjustments may be needed, for instance, different categories in the diagram (e.g., "hypothesis" elements in the science domain), different sentence openers (e.g., the opener "According to a statistic / estimate..." may or may not be appropriate for legal debates), or different time allocations for the specific phases.

Despite the reported success, the approach taken in this study has important limitations. Similar to other static collaboration scripts, the same support structure is provided to all students, irrespective of whether support is actually needed or not by specific students. This raises the following issues: (1) Constraints that are imposed on student collaboration may interfere rather than help ("over-scripting" [Dillenbourg 2002]). For instance, fixed-length activity phases do not account for different paces, that is, slower dyads might not have enough time to complete important activities; faster dyads might have to linger over activities they have already completed, wasting time that could be used fruitfully in other ways. (2) Problems that only concern a minority of students (e.g., specific types of sentence openers that may help only a minority of students) may not be addressed in the script. Consequently, some students might not get support that they would actually benefit from. (3) Some support is only useful in direct and immediate response to a problem. Static scripts, however, cannot diagnose problems online and react appropriately to students' needs or provide timely feedback. For instance, immediate feedback would be desirable when a discussion gets bogged down or stuck, or important sub-aspects have not been covered, with time running out in the process. (4) Adaptive fading of scaffolds (Pea 2004) is not possible since there are no means to detect whether scaffolding is no longer required.

Outlook

Looking ahead, we see possible technological enhancements to provide *adaptive* support to complement or support the static script. To improve the discussion quality beyond what can be achieved with a static script, we envision a software coach that uses both straightforward analysis techniques and Artificial Intelligence (AI) to provide adaptive guidance (Scheuer et al. 2012). For instance, in the example chat discussed in section "Case Analysis", both students only support their own arguments without addressing the points made by their partner, as witnessed by the excessive (and almost exclusive) use of proponent sentence openers. A relatively straightforward action a software coach could take would be to encourage critical comments on their partner's argument (e.g., "You are making many points in favor of your argument, but what about your partner's argument? Try to identify the weaknesses in your partner's position."). Technically, this kind of analysis could be implemented in a rather straightforward way, focusing on the number of specific sentence openers used, without the software understanding what is said. Notwithstanding, such feedback can be effective in promoting the discussion quality in terms of discourse coherence (since objections typically target previous statements and beg for a direct response) and elaboration of meaning (the response might clarify or modify a criticized statement). As another example of straightforward support, if a discussion bogs down (e.g., no contribution for *x* seconds), the software coach could initiate a new discussion topic, e.g., by asking canned, thought-provoking questions that may trigger student reasoning and reactions (Asterhan and Schwarz 2010).

A more challenging problem is to provide support on the content-level, which requires some level of machine understanding of language content. For instance, the software coach might suggest specific arguments to attack one of the partner's statements or might advise moving on to an important topic not yet covered. To provide such support the coach must accomplish two tasks: (a) understanding student contributions and (b) generating appropriate responses.

Starting with (b), since it is probably the easier of the two tasks for the software coach, an expert-created knowledge base could be utilized, which formally represents the "space of debate" for the given topic, including claims, arguments, and evidence potentially involved in a discussion about that topic. Such a knowledge base would empower the software coach to reason about the discussion domain, an important prerequisite to provide content-level feedback. Similar models have been used in the past to support argument-diagramming activities (e.g., Suthers et al. 2001; Dragon et al. 2006).

Point (a) is a thornier problem, since free-text messages must be classified and correlated with the coach's knowledge base. One possibility is the use of models and classifiers learned from a corpus of data (Rosé et al. 2008; McLaren et al. 2010; D'Mello et al. 2010). Alternatively, students could provide structured input that can be understood by a computer coach without sophisticated (but possible less reliable) computational models and techniques. For instance, students could provide *explicit* references and thereby reveal the contents they are talking about. Chat messages might reference diagram boxes, diagram boxes might reference passages of the given texts, and text passages might be annotated with knowledge base elements. Such references can provide "anchors" to correlate discussion, diagram and knowledge base contents. One crucial principle when attempting to have students make explicit references is that they must recognize and realize the benefits of making such references. Therefore, an important design objective is to avoid additional workload on the student, assuring user acceptance and avoiding interference with the learning task. In the study described in this paper students were instructed to – and actually did use – hash-tags in their chat messages to clarify which diagram boxes they were talking about (e.g., "I don't agree at all with the claim made in #12"), demonstrating the ease with which students can adopt such a referencing practice.

Conclusion

In conclusion, we have developed a collaborative learning environment, comprising an argument-diagramming tool and a collaboration script, to promote the learning of collaborative argumentation skills and content knowledge. In an empirical study, conducted in the domain of climate change ethics, we have shown that an additional proponent / critic role script has a positive influence on the discussion quality and students' perception of their learning. The environment is currently static, unresponsive to actions that students take while working with the software, and we believe this does not support the full potential of our approach. In the next phase of our research, we will extend the environment to adaptively support students, through a software coach that monitors student actions as they work and learn together and then provides appropriate feedback to help students collaborate better and, consequently, learn more.

Acknowledgements: We would like to thank Toby Dragon, Christoph Fehige, Vera Gehlen-Baum, Dimitra Tsovaltzi, and Florian Zickwolf for advice and support in planning, organizing and conducting the second study. This work was supported by the German Research Foundation under the grant "LASAD—Learning to Argue: Generalized Support Across Domains."

References

Andriessen, J. Baker, M., & Suthers, D. (2003). *Arguing to Learn: Confronting Cognitions in Computer-Supported Collaborative Learning Environments*. Dordrecht: Kluwer.

Andriessen, J. (2006). Arguing to Learn. In K. Sawyer (Ed.), *Handbook of the Learning Sciences* (pp. 443–459). New York: Cambridge University Press.

Andriessen, J., & Schwarz, B. (2009). Argumentative Design. In N. Muller Mirza, & A.-N. Perret-Clermont (Eds.), *Argumentation and Education: Theoretical Foundations and Practices* (pp. 145–174). New York: Springer.

Asterhan, C.S.C., & Schwarz, B.B. (2009). The role of argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialogue. *Cognitive Science, 33*(3), 373–400.

Asterhan, C.S.C., & Schwarz, B.B. (2010). Online Moderation of Synchronous e-Argumentation. *International Journal of Computer-Supported Collaborative Learning, 5*(3), 259–282.

Baker, M. J., & Lund, K. (1997). Promoting reflective interactions in a computer-supported collaborative learning environment. *Journal of Computer Assisted Learning*, *13*, 175–193.

Baker, M. (2003). Computer-mediated argumentative interactions for the co-elaboration of scientific notions. In J. Andriessen, M. Baker, & D. Suthers (Eds.), *Arguing to learn: confronting cognitions in computer-supported collaborative learning environments* (pp. 1–25). Dordrecht: Kluwer.

Baker, M., Andriessen, J., Lund, K., van Amelsvoort, M., & Quignard, M. (2007). Rainbow: A Framework for Analyzing Computer-Mediated Pedagogical Debates. *International Journal of Computer-Supported Collaborative Learning*, *2*(2–3), 247–272.

Bell, P. (1997). Using argument representations to make thinking visible for individuals and groups. In R. Hall, N. Miyake & N. Enyedy (Eds.), *Proceedings of CSCL-1997* (pp. 10–19). Toronto, Ontario, Canada: University of Toronto Press.

Buckingham Shum, S., MacLean, A., Bellotti, V. M. E., & Hammond, N. V. (1997). Graphical Argumentation and Design Cognition. *Human-Computer Interaction*, *12*(3), 267–300.

Cavalier, R., & Weber, K. (2002). Learning, Media, and the Case of Dax Cowart: A Comparison of Text, Film, and Interactive Media. *Interactive Learning Environments, 10*(3), 242–262.

Chaudhuri, S., Kumar, R., Howley, I., & Rosé, C. P. (2009). Engaging Collaborative Learners with Helping Agents. In V. Dimitrova, R. Mizoguchi, B. du Boulay, & A. Graesser (Eds.), *Proceedings of the 14th International Conference on Artificial Intelligence in Education (AIED 2009)* (pp. 365–372). Amsterdam: IOS Press.

Clark, D. B., D'Angelo, C. M., & Menekse, M. (2009). Initial Structuring of Online Discussions to Improve Learning and Argumentation: Incorporating Students' Own Explanations as Seed Comments Versus an Augmented-Preset Approach to Seeding Discussions. *Journal of Science Education and Technology, 18*(4), 321–333.

Dansereau, D. F. (2005). Node-Link Mapping Principles for Visualizing Knowledge and Information. In: S.-O. Tergan, & T. Keller (Eds.), *Knowledge and Information Visualization* (pp. 61–81), LNCS 3426. Berlin: Springer.

Dillenbourg, P. (2002). Over-scripting CSCL. In P. A. Kirschner (Ed.), *Three worlds of CSCL: Can we support CSCL* (pp. 61–91). Heerlen: Open University of the Netherlands.

Dillenbourg, P., & Hong, F. (2008). The mechanics of CSCL macro scripts. *International Journal of Computer Supported Collaborative Learning, 3*(1), 5–23.

D'Mello, S., Olney, A., & Person, N. (2010). Mining Collaborative Patterns in Tutorial Dialogues. *Journal of Educational Data Mining, 2*(1), 1–37.

Doise, W., & Mugny, W. (1984). The Social Development of the Intellect. Oxford: Pergamon.

Dragon, T., Woolf, B. P., Marshall, D., & Murray, T. (2006). Coaching Within a Domain Independent Inquiry Environment. In M. Ikeda, K.D. Ashley, & T.-W. Chan (Eds.), *Proceedings of the 8th International Conference on Intelligent Tutoring Systems, LNCS 4053* (pp. 144–153). Berlin: Springer.

Dragon, T., Mavrikis, M. McLaren, B. M., Harrer, A., Kynigos, C., Wegerif, R., & Yang, Y. (in press). Metafora: A web-based platform for learning to learn together in science and mathematics. To be published in *IEEE Transactions on Learning Technologies*.

Easterday, M. W., Aleven, V., Scheines, R., & Carver, S. M. (2009). Constructing Causal Diagrams to Learn Deliberation. *International Journal of Artificial Intelligence in Education, 19*(4), 425–445.

Erkens, G., Jaspers, J., Prangsma, M., & Kanselaar, G. (2005). Coordination processes in computer supported collaborative writing. *Computers in Human Behavior, 21*(3), 463–486.

Harrell, M. (2008). No Computer Program Required: Even Pencil-and-Paper Argument Mapping Improves Critical-Thinking Skills. *Teaching Philosophy, 31*(4), 351–374.

Janssen, J., Erkens, G., Kirschner, P., & Kanselaar, G. (2010). Effects of representational guidance during computer-supported collaborative learning. *Instructional Science, 38*(1), 59–88.

Jeong, A. & Joung, S. (2007). Scaffolding collaborative argumentation in asynchronous discussions with message constraints and message labels. *Computers & Education*, 48(3), 427–445.

Jermann, P., & Dillenbourg, P. (2003). Elaborating new Arguments through a CSCL Script. In J. Andriessen, M. Baker, & D. Suthers (Eds.), *Arguing to Learn: Confronting Cognitions in Computer-Supported Collaborative Learning Environments* (pp. 205–226). Dordrecht: Kluwer.

Johnson, D.W., & Johnson, R. T. (1991). *Learning together and alone*. *Englewood Cliffs*, NJ: Prentice Hall.

Johnson, D.W., & Johnson, R. T. (1994). Constructive conflict in the schools. *Journal of Social Issues, 50*(1), 117–137.

Johnson, D. W., & Johnson, R. T. (2009). An Educational Psychology Success Story: Social Interdependence Theory and Cooperative Learning. *Educational Researcher*, *38*(5), 365–379.

Kirschner, P. A., Buckingham Shum, S. J., & Carr, C. S. (Eds.). (2003). *Visualizing argumentation: software tools for collaborative and educational sense-making*. London: Springer.

Kuhn, D. (1991). *The Skills of Argument*. New York: Cambridge University Press.

Lazonder, A. W., Wilhelm, P., & Ootes, S. A. W. (2003). Using Sentence Openers to Foster Student Interaction in Computer-Mediated Learning Environments. *Computers & Education, 41(3)*, 291–308.

Loll, F., Pinkwart, N., Scheuer, O., & McLaren, B. M. (2012). How Tough Should It Be? Simplifying the Development of Argumentation Systems using a Configurable Platform. In N. Pinkwart, N., & B. M. McLaren (Eds.), *Educational Technologies for Teaching Argumentation Skills* (pp. 169–197). Bentham Science Publishers.

Loui, R. P., Norman, J., Altepeter, J., Pinkard, D., Linsday, J., & Foltz, M. (1997). Progress on room 5: A testbed for public interactive semi-formal legal argumentation. In *Proceedings of the 6th International Conference on Artificial Intelligence and Law (ICAIL 1997)* (pp. 207–214). ACM Press.

Lund, K., Molinari, G., Séjorné, A., & Baker, M. (2007). How do argumentation diagrams compare when student pairs use them as a means for debate or as a tool for representing debate? *International Journal of Computer-Supported Collaborative Learning, 2*(2–3), 273–295.

McAlister, S., Ravenscroft, A., & Scanlon, E. (2004). Combining Interaction and Context Design to Support Collaborative Argumentation Using a Tool for Synchronous CMC. *Journal of Computer Assisted Learning*, *20*(3), 194–204.

McLaren, B. M., Scheuer, O., & Mikšátko, J. (2010). Supporting collaborative learning and e-Discussions using artificial intelligence techniques. *International Journal of Artificial Intelligence in Education*, *20*(1), 1–46.

McManus, M. M., & Aiken, R. M. (1995). Monitoring computer-based problem solving. *Journal of Artificial Intelligence in Education, 6*(4), 307–336.

McManus, M. M., & Aiken, R. M. (1996). Teaching collaborative skills with a group leader computer tutor. *Education and Information Technologies, 1*, 75–96.

Munneke, L., van Amelsvoort, M., & Andriessen, J. (2003). The role of diagrams in collaborative argumentation-based learning. *International Journal of Educational Research, 39*(1–2), 113–131.

Noroozi, O., Weinberger, A., Biemans, H. J. A., Mulder, M., & Chizari, M. (2012). Argumentation-based computer supported collaborative learning (ABCSCL). A systematic review and synthesis of fifteen years of research. *Educational Research Review, 7*(2), 79–106.

Novak, J. D. (1990). Concept maps and vee diagrams: Two metacognitive tools for science and mathematics education. *Instructional Science, 19*(1), 29–52.

Nussbaum, E. M., Hartley K., Sinatra, G. M., Reynolds, R. E., & Bendixen, L. D. (2002). Enhancing the Quality of On-Line Discussions. Paper presented at the *Annual Meeting of the American Educational Research Association*.

Nussbaum, E. M. (2008). Collaborative Discourse, Argumentation, and Learning: Preface and Literature Review. *Contemporary Educational Psychology*, *33*(3), 345–359.

O'Donnell, A. M., Dansereau, D. F., & Hall, R. H. (2002). Knowledge Maps as Scaffolds for Cognitive Processing. *Educational Psychology Review*, *14*(1), 71–86.

Oh, S. & Jonassen, D. H. (2007). Scaffolding online argumentation during problem solving. *Journal of Computer Assisted Learning, 23*(2), 95–110.

Pea, R. D. (2004). The Social and Technological Dimensions of Scaffolding and Related Theoretical Concepts for Learning, Education, and Human Activity. *The Journal of the Learning Sciences, 13*(3), 423–451.

Pinkwart, N., Ashley, K. D., Lynch, C., & Aleven, V. (2009). Evaluating an Intelligent Tutoring System for Making Legal Arguments with Hypotheticals. *International Journal of Artificial Intelligence in Education, 19*(4), 401–424.

Postmes, T., Spears, R., & Cihangir, S. (2001). Quality of Decision Making and Group Norms. *Journal of Personality and Social Psychology*, *80*(6), 918–930.

Ravenscroft, A. (2007). Promoting Thinking and Conceptual Change with Digital Dialogue Games. *Journal of Computer Assisted Learning, 23*, 453–465.

Rittel, H. W. T., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*, 155–169.

Robertson, J., Good, J., & Pain, H. (1998). BetterBlether: The Design and Evaluation of a Discussion Tool for education. *International Journal of Artificial Intelligence in Education 9*(2), 219–236.

Rosé, C., Wang, Y.-C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing Collaborative Learning Processes Automatically: Exploiting the Advances of Computational Linguistics in Computer-Supported Collaborative Learning. *International Journal of Computer-Supported Collaborative Learning 3*(3), 237–271.

Rummel, N., & Spada, H. (2005). Learning to Collaborate: An Instructional Approach to Promoting Collaborative Problem Solving in Computer-Mediated Settings. *Journal of the Learning Sciences*, *14*(2), 201–241.

Scardamalia, M., & Bereiter, C. (2006). Knowledge Building: Theory, Pedagogy, and Technology. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 97–118). New York: Cambridge University Press.

Scheuer, O., Loll, F., Pinkwart, N., & McLaren, B. M. (2010). Computer-Supported Argumentation: A Review of the State of the Art. *International Journal of Computer-Supported Collaborative Learning, 5*(1), 43–102.

Scheuer, O, McLaren, B. M., Harrell, M., & Weinberger, A. (2011). Scripting Collaboration: What Affects Does it Have on Student Argumentation? In T. Hirashima et al. (Eds.), *Proceedings of the International Conference on Computers in Education (ICCE-2011)* (pp. 181–188). Asia-Pacific Soc. for Computers in Education.

Scheuer, O., McLaren, B. M., Loll, F. & Pinkwart, N. (2012). Automated Analysis and Feedback Techniques to Support and Teach Argumentation: A Survey. In N. Pinkwart, & B. M. McLaren (Eds.), *Educational Technologies for Teaching Argumentation Skills* (pp. 71–124). Bentham Science Publishers.

Scheuer, O., & McLaren, B. M. (2013). CASE: A Configurable Argumentation Support Engine. *IEEE Transactions on Learning Technologies*. doi:10.1109/TLT.2013.3

Schwarz, B. B., & Glassner, A. (2007). The role of floor control and of ontology in argumentative activities with discussion-based tools. *International Journal of Computer Supported Collaborative Learning, 2*(4), 449–478

Soller, A. (2001). Supporting social interaction in an intelligent collaborative learning system. *International Journal of Artificial Intelligence in Education, 12*, 40–62.

Stegmann, K., Weinberger, A., & Fischer, F. (2007). Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning*, *2*(4), 421–447.

Stegmann, K., Wecker, C., Weinberger, A., & Fischer, F. (2012). Collaborative argumentation and cognitive elaboration in a computer-supported collaborative learning environment. *Instructional Science, 40*(2), 297–323.

Suthers, D. D., Weiner, A., Connelly, J., & Paolucci, M. (1995). Belvedere: Engaging students in critical discussion of science and public policy issues. In J. Greer (Ed.), *Proceedings of the 7th World Conference on Artificial Intelligence in Education (AI-ED 1995)* (pp. 266–273). Charlottesville, VA: Association for the Advancement of Computing in Education.

Suthers, D. D., Connelly, J., Lesgold, A., Paolucci, M., Toth, E., Toth, J., & Weiner, A. (2001). Representational and Advisory Guidance for Students Learning Scientific Inquiry. In K. D. Forbus, & P. J. Feltovich, (Eds.), *Smart machines in education: The coming revolution in educational technology* (pp. 7–35). AAAI/MIT Press.

Suthers, D. (2003). Representational Guidance for Collaborative Inquiry. In J. Andriessen, M. Baker, & D. Suthers (Eds.), *Arguing to Learn: Confronting Cognitions in Computer-Supported Collaborative Learning Environments* (pp. 27–46). Dordrecht: Kluwer.

Suthers, D. D., & Hundhausen, C. (2003). An experimental study of the effects of representational guidance on collaborative learning processes. *Journal of the Learning Sciences, 12*(2), 183–219.

Suthers, D. D., Vatrapu, R., Medina, R., Joseph, S., & Dwyer, N. (2008). Beyond threaded discussion: Representational guidance in asynchronous collaborative learning environments. *Computers & Education, 50*(4), 1103–1127.

Toulmin, S. E. (1958). *The Uses of Argument*. New York: Cambridge University Press.

Twardy, C. R. (2004). Argument Maps Improve Critical Thinking. *Teaching Philosophy, 27*(2), 95–116.

Van Eemeren, F. H., & Grootendorst, R. (2004). *A Systematic Theory of Argumentation: The Pragma-Dialectical Approach*. Cambridge: Cambridge University Press.

van Gelder, T. (2002). Argument Mapping with Reason!Able. *The American Philosophical Association Newsletter on Philosophy and Computers*, *2*(1), 85–90.

van Gelder, T. (2005). Teaching critical thinking: some lessons from cognitive science. *College Teaching, 53*, 41–46.

Wegerif, R., McLaren, B. M., Chamrada, M., Scheuer, O., Mansour, N., Mikšátko, J., & Williams, M. (2010). Exploring Creative Thinking in Graphically Mediated Synchronous Dialogues. *Computers & Education, 54*(3), 613–621.

Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer–supported collaborative learning. *Instructional Science*, *33*(1), 1–30.

Weinberger, A., & Fischer, F. (2006). A Framework to Analyze Argumentative Knowledge Construction in Computer-Supported Collaborative Learning. *Computers & Education, 46*(1), 71–95.

Weinberger, A., Stegmann, K., & Fischer, F. (2007). Knowledge convergence in collaborative learning: Concepts and assessment. *Learning and Instruction*, *17*(4)**,** 416–426.

Weinberger, A., Stegmann, K., & Fischer, F. (2010). Learning to Argue Online: Scripted Groups Surpass Individuals (Unscripted Groups do not). *Computers in Human Behavior*, *26*(4), 506–515.

Wigmore, J. H. (1931). *The principles of judicial proof (2nd ed.)*. Boston: Little, Brown & Co.