CrossMark

How to improve argumentation comprehension in university students: experimental test of a training approach

Sarah von der Mühlen[1] · Tobias Richter[1] · Sebastian Schmid[2] · Kirsten Berthold[3]

Abstract

The ability to comprehend informal arguments is essential for scientific literacy but students often lack structural knowledge about these arguments, especially when the arguments are more complex. This study used a pre-post-test design with a follow-up 4 weeks later to investigate whether a computerised training in identifying structural components of informal arguments can improve university students' competences to understand complex arguments. The training was embedded in a constructivist learning environment and contents were based on the Toulmin model of argument structure, according to which arguments can be deconstructed into several functional components: Claim, datum, warrant, backing evidence, and rebuttal. Being able to identify the warrant is central for scientific literacy, as the warrant determines whether a conclusion is justified given the data. Results indicate that training in argument structure did not generally improve performance for all students and argument types, but that it was particularly helpful for identifying more complex arguments with a less typical structure and relational aspects between key components (i.e. warrants). High achieving students profited the most from this intervention, and the intervention was also helpful for students with high pretest accuracy scores. Our results suggest that interventions to foster argumentation skills should be included into the curriculum and these interventions should be designed to match learners' ability level.

Keywords Argument comprehension · Epistemic competences · Informal arguments · Training

✉ Tobias Richter
tobias.richter@uni-wuerzburg.de

Sarah von der Mühlen
sarah.von-der-muehlen@uni-wuerzburg.de

Sebastian Schmid
seb.schmid@paedagogik.uni-regensburg.de

Kirsten Berthold
kirsten.berthold@uni-bielefeld.de

[1]   Department of Psychology IV, University of Würzburg, Röntgenring 10, 97070 Würzburg, Germany

[2]   Department of Educational Science, University of Regensburg, Universitätsstraße 31, 93053 Regensburg, Germany

[3]   Department of Psychology, Bielefeld University, Postfach 10 01 31, 33501 Bielefeld, Germany

Argument comprehension skills are essential for learning and decision-making across the lifespan. Lay people interested in socioscientific issues such as risks of cell phones, media, vaccinations, or genetically modified food (cf. Sadler 2004) are confronted with an overwhelming number of different, and often conflicting, arguments. Similarly, when university students learn about a scientific topic, they are required to read a variety of documents, many of which contain opposing evidence for different theoretical claims. Being able to comprehend the claims and arguments presented in different texts comprises an essential aspect of scientific literacy (Britt et al. 2014). Knowledge about how an argument is structured is essential for understanding scientific information and for determining the quality of an argument (Britt et al. 2014; Britt and Larson 2003; Wolfe et al. 2009).

Nevertheless, a considerable number of students possess insufficient skills to comprehend arguments (e.g., National Assessment of Educational Progress 1996; OECD 2011, 2014). For example, results from the Programm for International Student Assessment (PISA) for reading and scientific literacy revealed that the majority of high school students were able to use basic scientific knowledge to identify a valid conclusion or scientific evidence for a claim, but only a minority of them were able to identify more complex arguments, to use evidence for evaluating the quality of arguments, link different knowledge, or apply relevant knowledge to unfamiliar or real-life situations (OECD 2014). Similarly, only a small number of students were able to discriminate between relevant and irrelevant information. Although German students performed slightly higher than the OECD average for scientific literacy, these students faced similar problems.

The present research investigated the effects of a training intervention designed to improve students' competences to comprehend more complex informal arguments in scientific discourse–arguments that students typically encounter in the course of their studies. We begin with an analysis of the skills required to understand such arguments. In this context, we outline the Toulmin model of argumentation (Toulmin 1958) to describe the typical structure of an argument. Following this, we discuss frequent challenges that students face when trying to comprehend informal arguments and the conditions under which training in argumentation might be effective for overcoming these challenges (e.g., Hefter et al. 2014, 2015; Larson et al. 2004). We then present results from an argumentation training intervention based on Jonassen's (1999) constructivist learning environment approach. The experiment aimed at improving students' familiarity with the structure of informal arguments by teaching them how to identify different argument components and their relations.

Understanding informal arguments

Scientific texts are often structured like arguments, stating different (usually empirical) evidence for theoretical claims, including counter-arguments and limitations of the evidence. To understand such texts, readers construct a *mental model* of the situation described in the text from their general prior knowledge, i.e. a referential representation of the arguments' content (Johnson-Laird 1983). This mental model represents the state of affairs described in the message rather than the message itself. It helps the reader to establish connections between ideas within the text and between ideas stated in the text and prior knowledge about the content of the text (Chi et al. 1989). Thus, forming an accurate mental model is essential for a deeper understanding of the information presented in a text (Mayer 1989).

An argument is an attempt to convince the reader to accept a proposition, or claim (Galotti 1989). Arguments found in empirical scientific documents are often informal,

rather than formal arguments, and their quality cannot be determined by formal, deductive logic (Galotti 1989; Toulmin 1958). In a formal deductive argument, the conclusion follows with logical necessity from the premises. Formal arguments are truth-preserving in that the conclusion is necessarily true provided that the premises are true, Instead, in a strong, informal argument, the conclusion *probably* follows from the stated evidence (Voss and Means 1991; Voss et al. 1991). Scientific claims are often not certain facts, but interpretations of (usually empirical) evidence that are open for criticism and can be challenged with new information (e.g., by presenting counterevidence). Although, similar to formal arguments, informal arguments consist of a claim and one or more reasons, they may contain additional components. Toulmin (1958) proposed his model in reaction to the traditional formal reasoning perspective. According to Toulmin's argumentation model, full-fledged arguments contain a number of functional key components: a claim, reason(s) (or datum/data), a warrant, backing evidence, and a rebuttal (Toulmin 1958). The *claim* is the main statement being argued for. Claims are, by definition, controversial, and need to be supported with theoretical or empirical evidence which is referred to as *datum* (or *data)*. Claims and data are connected by the *warrant*. The warrant determines the strength of the evidence for the main claim, or, in other words, indicates whether the conclusion can be justified given the data. Another component, called *backing evidence,* provides (empirical or theoretical) support for the warrant. Finally, *rebuttals* contain counter-arguments or indicate circumstances in which the argument does not hold true.

Consider the following example (a brief summary of a study by Freeman et al. 2017):

People should not eat eggs (*claim*), because eggs contain high amounts of cholesterol (*datum*). High amounts of cholesterol are unhealthy (*warrant*), because they may lead to coronary diseases (*backing*). However, individual factors play an important role and eggs may not increase the risk for coronary diseases in all people (*rebuttal*).

The claim that people should not eat eggs is supported by the datum that eggs contain high amounts of cholesterol. The datum lends support to the claim only on account of the warrant that high amounts of cholesterol are unhealthy. Backing for the warrant is stated by referring to the finding that high amounts of cholesterol may lead to coronary diseases. However, the argument does not apply to all people, but individual factors play an important role. This last sentence constitutes the rebuttal.

Scientific texts are often structured like full-fledged arguments (Suppe 1998). They present data (i.e. reasons), show the relevance of the observations to a scientific problem, provide a detailed description of data collection and analysis methods, justify their claims and interpretations of the evidence, and generate alternative explanations. Whereas the warrant and its corresponding backing evidence are often not explicitly stated in everyday arguments, but need to be inferred by the reader (e.g., Chambliss 1995), in a scientific texts, it is crucial to explicitly state why a particular conclusion is drawn from the results. Thus, warrants are particularly important in the scientific domain.

Typically, the order in which the different components are presented is hierarchical, whereby the claim holds the top position because all other components are presented to either support or oppose the main claim (*claim-first arguments*, Britt and Larson 2003). However, arguments can also be stated in a less typical way. For example, they can begin with the datum, followed by the main claim (*reason-first arguments*), or with the rebuttal (e.g., Larson et al. 2004). Typical arguments are processed faster and more accurately than less typical arguments, because they are usually more congruent with the readers' current mental model (Britt and Larson 2003). Most arguments contain linguistic markers or connectives like "therefore" or "because". These markers provide

important processing or conceptual information, because they signal relations across the different components, thereby helping the reader to construct a coherent representation of the text. Britt and Larson (2003) found that arguments with markers are processed faster than arguments without these signals and that statements including modal verbs (e.g., *should*) and uncertainty markers (e.g., *probably*) signaled controversial statements requiring support.

Awareness of an (accurate) argument schema, including relevant markers (e.g., modals and qualifiers), can help the reader to identify the main claim, link the data to this claim, guide coherence inferences, activate possible alternative explanations, and form a corresponding mental model—cognitive processes that are not only relevant for the comprehension, but also for the evaluation of informal arguments (Shaw 1996).

The challenges of dealing with informal arguments among lay readers

A number of studies suggest that lay readers use epistemic reasoning skills (i.e. skills that relate to the ability to form a valid understanding of a text; e.g., Richter 2011) to guide comprehension of arguments to some extent (see Johnson et al. 2004, for a review), but that they are not always accurate in doing so. Even younger students seem to use argument schemas to guide comprehension if the structure of arguments is made explicit to them (e.g., Chambliss 1995; Chambliss and Murphy 2002). For example, Chambliss (1995) provided high school students (12th graders) with clearly structured argumentative texts that included strong syntactical elements (signals) and introductory and concluding paragraphs that summarised the structure of the text. She found that students were able to recognise the argument structure and signalling text cues, and used them to guide comprehension and to construct accurate representations of the argument. However, Larson et al. (2004) noted that given their optimized structure, the arguments used in Chambliss's (1995) study were rather atypical for informal arguments, and did not reflect the complexity of authentic arguments. In their study, Larson et al. (2004) used a variety of more authentic arguments that included arguments with a less typical structure and found that university students identified only 30% of their key components correctly. For example, the students in their experiment often misidentified uncontroversial and unsupported statements, data, and even counter-arguments (when the rebuttal was stated first) as the main claim. Similarly, von der Mühlen et al. (2016) used think-aloud protocols to compare the performance of experts (advanced doctoral and post-doctoral students) with that of introductory university students and found that undergraduates struggled to identify key components of the Toulmin model, especially warrants.

Further evidence suggests that students seem to have particular difficulties to adequately represent relations between argument components, and that these problems are related to their difficulties to evaluate the quality of arguments (e.g., Britt and Kurby 2005; Larson et al. 2009; Shaw 1996). One explanation may be that evaluations of relational aspects between argument components are more effortful (Shaw 1996). Readers need to access relevant prior knowledge from memory, activate alternative explanations, and keep this information activated in working memory.

Thus, lay readers often seem to struggle with the comprehension (and evaluation) of more complex arguments. They lack relevant structural knowledge and find it particularly difficult to attend to relations between argument components, such as warrants.

Improving lay readers' competences to comprehend informal arguments in constructivist learning environments

The difficulties among students to correctly comprehend arguments highlight the need for explicit instruction and training of the strategies involved. Part of the problem may be that students have never received formal training in the skill of argumentation (Perkins 1985). Although students entering university are expected to possess relevant argumentation skills, they usually have little experience with more complex arguments. Textbooks, the dominant genre type used in high school classrooms, rarely contain complex arguments (Calfee and Chambliss 1988; Paxton 1997) and underlying relationships are often neglected (Beck 1989). Past research indicates that students may require practice in understanding the connection between data and claim (e.g., Larson et al. 2004, 2009; Shaw 1996; von der Mühlen et al. 2016).

Constructivist learning environments (*CLE*, Jonassen 1999), which are based on the assumption that knowledge cannot be transmitted but is individually constructed by the learner, have been shown to be effective for instruction in a number of interventions (e.g., Berthold and Renkl 2010; Hefter et al. 2014, 2015; Larson et al. 2009). Various research shows that students remember information better when they construct their own knowledge (e.g., De Winstanley and Bjork 2004; Marsh et al. 2001). This type of learning helps to achieve a deeper understanding of the material (Chi et al. 1989). Interactive environments, in which learners are allowed to correct their responses and in which information is easily accessible, are helpful elements of a CLE (Jonassen 1999). For example, in an intervention that aimed at improving students' evaluation of informal arguments, Larson et al. (2009) used individual knowledge construction (i.e. an interactive text) as a central element for instruction and successfully improved students' understanding of arguments. To this end, they focused on helping students to actively represent different components of two-clause (claim, reason) arguments.

The use of learning goals for authentic real-world problems and immediate feedback are also central elements of a CLE (Jonassen 1999). Informative feedback is crucial, because it increases motivation (Deci 1971), helping the learner to deeply process information (Jonassen 1999). For example, Larson et al. (2004) attempted to improve students' understanding of complex arguments, including arguments with a less typical structure, by developing a short (10 min) tutorial in which they defined key components of arguments and named a number of steps for comprehending arguments (e.g., writing down the main claim and supporting data). They also included linguistic markers to signal relationships between argument components and to help the reader to appropriately connect the various argument components into a coherent structure. Larson et al. (2004) showed that teaching the structure of an argument helped students to shift their attention towards relations between argument components when immediate feedback was provided. Moreover, they showed that students should be instructed with a clear learning goal that matches the task, as the tutorial was only successful when the goal was to comprehend (but not evaluate) the argument.

In addition, varied examples or cases of a problem should be included to represent complexity and enable cognitive flexibility. Experts can serve as *cognitive models* who demonstrate different cases (examples) of the problem and relevant strategies required to solve the problem (Jonassen 1999; Renkl 2009). Such illustrations can reduce cognitive complexity and help the learner to deeply process information during the practice phase (Renkl 2009). Video tutorials are particularly useful, because they stimulate both visual and auditory channels and thereby reduce cognitive complexity (Mousavi et al. 1995).

Finally, instructional prompts, in which learners are required to self-explain stated information, have been shown particularly useful for the acquisition of knowledge, because they stimulate deep processing of information (Berthold and Renkl 2010).

Thus, it appears that even short-term interventions, with a focus on providing knowledge about structural components of arguments and their relations, including linguistic markers, immediate feedback, and clear learning goals, can be a promising approach to help students improve their argument comprehension skills (Larson et al. 2004). Moreover, a variety of authentic argument types should be used for such an intervention.

The present research

The present research investigated whether training in argument structure can improve psychology students' competences to comprehend informal arguments. In particular, it was examined which types of arguments are particularly challenging and whether students in the experimental condition would improve their performances to recognise different components of arguments, including arguments with a less typical structure and those with less typical components (i.e. warrants, backing). Furthermore, it was examined whether pretest accuracies would influence (or possibly moderate) performance, and whether students with higher average grades would profit more from such an intervention than others. In addition, it was investigated whether students who received the training in argument structure would feel more confident with the Toulmin model after the intervention. Finally, it was examined which parts of the learning environment would be perceived as most helpful by the participants.

Some argument components (i.e. warrants, backing evidence) are often not explicitly stated in a text and might therefore be more difficult to identify than other components, such as claims, reasons, or rebuttals (von der Mühlen et al. 2016). Assuming that argument comprehension requires abstract representations of the functional components of arguments and their interrelations (Britt et al. 2014; Britt and Larson 2003; Wolfe et al. 2009), it might be necessary to include more complex arguments in an intervention. Earlier research mainly focused on improving students' understanding of relatively simple claim-reason arguments (e.g., Britt et al. 2008; Britt and Larson 2003; Larson et al. 2009). In contrast to this research, teaching students to attend to relational aspects of argument components and using a variety of argument types, including arguments with a more typical and arguments with a less typical structure, as well as a combination of more difficult and less difficult arguments (i.e. arguments with and without explicitly stated warrants), was a major concern of the present research. In addition, we included linguistic markers in our intervention to help students represent different argument components and signal relationships between these components. Britt and Larson (2003) found that students are able to use such markers to identify claims.

The study also extends prior research by considering characteristics of the reader. We were particularly interested in a possible (moderating) influence of study performance on the effects of our training intervention. Assuming that students with better study performance are more likely to be familiar with a broad range of scientific texts, this might (implicitly) provide them with some relevant prior knowledge (i.e. *discipline expertise*, Rouet et al. 1997) about the structure of arguments. This structural prior knowledge, in turn, should allow them to more easily integrate and apply information from the training intervention. Furthermore, their familiarity with various scientific texts might generally

foster their ability to understand informal arguments (Britt et al. 2014; Rouet et al. 1997). Similarly, we examined whether students with a certain level of understanding and/or familiarity with arguments in the pretest would profit (or not) from our intervention. Both study performance and pretest accuracies might involve (different kinds of) prior knowledge, or one or the other might reflect a more general cognitive ability, such as individual differences in the ability to think rationally (Stanovich 2012).

As knowledge about the structure of arguments has been shown to be particularly important for comprehension and evaluation (Britt et al. 2014; Britt and Larson 2003; Larson et al. 2004, 2009; Wolfe et al. 2009), the experiment evaluated the effects of an intervention designed to improve students' competences to recognise the structural components of informal arguments and their relations, including relevant markers. Building on earlier research (e.g., Hefter et al. 2014, 2015; Larson et al. 2009), our intervention conveyed both conceptual and procedural knowledge in a constructivist learning environment (Jonassen 1999). *Procedural knowledge* can be defined as the ability to execute sequences of action to solve problems (e.g., Rittle-Johnson et al. 2009), whereas *conceptual knowledge* can be defined as an integrated and functional understanding of domain-specific ideas (Kilpatrick et al. 2001; Rittle-Johnson et al. 2009). Both procedural and knowledge construction are important for learning to be effective, because they seem to develop iteratively (Rittle-Johnson et al. 2009). Conceptual knowledge is necessary for an accurate construction and execution of problem-solving procedures. Practice using such procedures, in turn, helps students to deepen their understanding of relevant concepts.

The training intervention was designed to increase students' familiarity with the basic structure of informal arguments and to improve their ability to recognise different components and their relations using the Toulmin (1958) model. In addition, it was examined whether study performance would influence or moderate posttest and follow-up accuracies. Finally, we investigated as an exploratory research question whether pretest accuracies would predict or moderate performance in the posttest and follow-up. The following research questions and hypotheses were formulated:

1. *How does argument structure affect argument comprehension?* We expected that arguments including less typical components (i.e. warrants, backing) and arguments with a less typical structure will be more challenging to identify than arguments with more typical components (i.e. claim, datum, rebuttal) and arguments with a typical, claim-first structure, as reflected in lower pretest accuracy scores for these arguments (Hypothesis 1).

2. *Does the argument structure training improve argument comprehension?* We expected that participants in the experimental condition will improve their comprehension of different components of the Toulmin model, as reflected in higher posttest and follow-up accuracy scores, compared to the control condition. In particular, we expect an improved performance for arguments with an atypical structure and less typical components (i.e. warrants and backing evidence; Hypothesis 2).

3. *Does the effectiveness of the argument struture training depend on students' prior abilities in argument comprehension?* Given the complexity of the text materials and the training, it is possible that students with relatively high performance already in the pretest will profit more from the training than low-performing students. However, given the lack of previous research on this issue, the (moderating) influence of pretest accuracies was examined as an exploratory research question (Research Question 1).

4. *Does the effectiveness of the argument structure training depend on students' prior abilities in argument comprehension?* In the same vein, it seems possible that the average grades of the participating students moderate the training effectiveness. Again, given the complexity of the text materials and the training, high-achieving students might particularly profit from the argument structure intervention, as reflected in higher post-test and follow-up accuracy scores, compared to the control condition. Given the lack of previous research, this was also examined as an exploratory research question (Research Question 2).

5. *Does the argument structure training increase students' confidence in identifying argument components?* We expected that participants in the experimental condition will feel more confident in dealing with the Toulmin model after the intervention (Hypothesis 3).

6. *Are the elements of the constructivist learning environment perceived as helpful?* In addition, we examined as an exploratory research question whether different elements from our constructivist learning environment (e.g., video-based tutorials, practical exercises, feedback, prompts) will be perceived as helpful by the participants in the experimental group (Research Question 3).

Participants in the control condition worked on a computerised speed-reading training, for which no effects on the ability to recognise argument components were expected. A training aimed at fostering speed-reading competences was chosen because this competence is very different from the competence to develop a deep understanding of arguments. In the speed-reading training, students focused on developing effective strategies to quickly locate and recognise important information from a text. To test the hypotheses, students' performance on a computerised pretest regarding the ability to identify different functional components of the Toulmin model was compared to the performance in a posttest and follow-up 4 weeks after the posttest, and to the performance of a control group who received a speed-reading training.

Method

Participants

Fifty-three psychology students (10 males, 43 females) with an average age of 24 years ($SD = 5.70$) participated in the study. The majority of students (37) were undergraduates in their second semester, nine of them were in their fourth semester, and four students were in their sixth semester. Three participants had started a Master's programme. Participants provided informed consent at the beginning of the experiment and were reimbursed with course credits or financial remuneration (8 Euros per hour) after the completion of all sessions. In addition, they received an optional feedback with regard to their progress a few weeks later.

Text materials

All materials were presented in German. The examples stated in the present paper were translated into English.

Text materials for the pretest, posttest, and follow-up

The text materials provided for the identification of different argument components were short argumentative texts with a mean length of 89 words in each argument. Three parallel versions were created based on von der Mühlen et al.'s (2016) study, and additional arguments were taken from their pilot study. The texts were summaries of existing empirical articles from different fields within the domain of psychology, adapted to fit the structure of Toulmin's (1958) model. Each of the versions contained four texts and one practice example. Three of those texts were full-fledged arguments, including a claim, a datum, a warrant, backing evidence, and a rebuttal (Toulmin 1958), and one of them contained only a claim, a datum, and a rebuttal. Two of the texts (including the argument consisting of three components) exhibited a typical structure (claim-first arguments, Britt and Larson 2003). The two remaining texts were atypically structured (reason-first arguments, Britt and Larson 2003). The texts had rather low readability scores that were representative of the literature students typically read ($M = 17$) as indexed by the German adaptation of Flesch's Reading Ease Index (Amstad 1978).

Text materials for the trainings

The training intervention conveyed both conceptual and procedural knowledge in a constructivist learning environment, using a cognitive modelling approach (Jonassen 1999). For our procedural knowledge measure, we included a variety of typical and less typical, as well as complex and less complex arguments to test whether participants could apply their knowledge to new and unfamiliar situations. The conceptual knowledge included information about the Toulmin model and linguistic markers and was given to students first to help them apply their knowledge about the structure of arguments and recognize different argument components in the practical phase. A theoretical introduction provided appropriate background knowledge about the structure of full-fledged arguments (Toulmin 1958). Learning goals and prompts were used to foster focused processing of the instructions, the central concepts of the explanations, and the practice items (Berthold and Renkl 2010). Based on Jonassen's (1999) cognitive modelling approach, two video tutorials were used to explain the strategies needed to correctly identify different argument components. In the practical part, participants worked on a number of argumentative texts. Feedback was provided for each task and participants were able to access relevant information (e.g., theoretical information, video tutorials, notes) when needed at all stages of the experiment (cf. Hefter et al. 2014, 2015).

Theoretical introduction

In the theoretical introduction, relevant knowledge about the structure, relevance, and purpose of informal arguments for scientific literacy (Britt et al. 2014; Britt and Larson 2003; Wolfe et al. 2009) was provided. The Toulmin (1958) model was explained using a visual scheme. All theoretical input was explained with several examples to reduce cognitive complexity (Mousavi et al. 1995) and enable deep cognitive processing during the practical exercises (Renkl 2009). These examples portrayed the problem (e.g., *Identify the claim of an argument*), pointed out different strategies to solve the problem (e.g., *Pay attention to markers*), and revealed the solution to the problem (e.g., *The first sentence is the claim*).

Furthermore, attention to markers of epistemic modality, such as *should*, and connectors such as *as a result,* or *therefore* (Britt and Larson 2003), was introduced as a strategy to recognise different argument components and their relations. A number of learning goals were formulated to foster focused processing of information (Berthold and Renkl 2010). These included three questions: (a) *What does the basic structure of an argument look like?* (b) *Which components does an argument include?* and (c) *How can we identify different argument components?*. Participants were prompted to answer these questions at different stages of the experiment.

Explanation prompts

Specific prompts requested participants to reproduce conceptual information. The following prompts were integrated into the learning environment: (a) *Name each argument component and enter them in a text field*; (b) *Assign each argument component to its corresponding position within the scheme using a dropdown button*; (c) *Provide a written definition of each component*; and (d) *Name useful strategies for recognising the components in an argument and write them down in a text field*.

Video tutorials

Two video tutorials were developed to convey the strategies needed to identify the components of arguments. Both tutorials included one full-fledged argument including a claim, a datum, a warrant, backing evidence, and a rebuttal (Toulmin 1958). Again, the arguments were summaries of existing empirical articles from different fields within the domain of psychology, adapted to fit Toulmin's (1958) model. The first tutorial (length: 03:41 min) described a typical argument (73 words), beginning with a claim and followed by the datum, the warrant, backing for this warrant, and a rebuttal. The second tutorial (length: 03:58 min) included an atypical argument (76 words) and began with the datum, followed by the warrant, backing for the warrant, the claim, and the rebuttal. A male model, who was portrayed as an expert in argumentation, read aloud both arguments and explained its structure in a stepwise fashion (cp. Jonassen 1999; Renkl 2009). Each argument component was explained separately and elaborative information was provided with an example. Markers signalling relations between argument components were highlighted in each statement and explained by the model. The two arguments can be found in "Appendix". Again, readability of the arguments was rather difficult ($M = 29$ in "Argument 1" vs. $M = 19$ in "Argument 2"), as indexed by the German adaptation of Flesch's Reading Ease Index (Amstad 1978).

Practice texts

The practice texts included 12 arguments. As in the video tutorials, the texts were based on existing empirical articles from different fields within the domain of psychology, summarised to represent each component of the Toulmin (1958) model. Generally, the structure of the arguments resembled the texts used in the pretest, posttest, and follow-up. Furthermore, different types of arguments were included to increase complexity (Jonassen 1993; Larson et al. 2004). The texts in the training included both full-fledged arguments (Toulmin 1958) and arguments with only three components (claim, datum, rebuttal), and both typical (claim-first) and atypical (reason-first) arguments (Britt and Larson 2003). The arguments had a mean length of 88 words. As in the tests and tutorials, the texts had rather low mean

readability scores ($M = 31$), as indexed by the German adaptation of Flesch's Reading Ease Index (Amstad 1978). However, readability was slightly higher due to the inclusion of simpler arguments with only three components.

Feedback

In every exercise, participants received immediate feedback on the correctness of their response, including the correct solution. In addition, a table showing general progress was provided. This table gave informative feedback on the number and types of argument components that had been assigned (in)correctly so that participants could repeat more difficult tasks.

Speed-reading training

For the control group, the application *Schneller Lesen* (reading faster, Heku-IT) was used to practice fast reading. The application consists of several exercises, whereby each takes about 60 s. These exercises are embedded in several superordinate lessons, each containing eight exercises. The most important strategies used by the application to improve speed-reading competences are avoiding setbacks, not reading every single word of a text silently to oneself, and conceiving groups of words as an entity to derive meaning. The application provides feedback by granting points for successfully completed exercises.

Validation of text and item materials

The text materials for the pretest, posttest, and follow-up were normed and validated in a study by Schroeder et al. (2008) and in the pilot study preceding the study by von der Mühlen et al. (2016). The correlation between parallel versions in this study was $r = .86$, $p < .01$.

For the argument structure training (i.e. the texts used in the tutorial and practice session), interrater reliability was determined by two doctoral candidates in the domain of psychology. There was high agreement among raters that all argument components in the training material were described and assigned correctly, Cohen's $\alpha = .95$. The speed-reading application has been tested and rated as "best product" by the leading German product testing group (Stiftung Warentest 2015) for improving speed-reading competences up to 50% and remembrance of a text without any decline in understanding.

Software

The testing software used to display the tests and to record responses and response times was Inquisit 3.0.6.0. It was run on four identical HP notebooks with 15″ screens. For the speed-reading training, Android OS, v4.4.2 (KitKat) was used for each of five identical ASUS tablet computers (10.1″) on which the application was installed.

Procedure

Participants were tested in groups of up to four people in a laboratory, and completed a total of four sessions, including a pretest, a training intervention, a posttest, and a follow-up. The interval between the pretest and the training intervention was 1 week, the posttest was conducted 15 min after the training session, and the follow-up was performed 4 weeks

later. Although participants were allowed as much times as they needed to complete the tasks, on average, the pretest took about 1 h, the combined training and posttest session approximately 90 min (60 min for the training and 30 min for the posttest), and the follow-up about 40 min. Apart from the argument structure test or speed-reading test, participants completed another task (i.e. evaluating the plausibility of arguments) in the pretest, post-test, and follow-up session, which will not be discussed in the present work.

Pretest

Upon arrival, participants were welcomed, briefly informed about the procedure, and seated in front of a computer where they gave informed consent to participate in the experiment. Study performance was assessed with self-reported average grades in their present course of studies. Subsequently, the participants worked on the argument structure test. The other parallel version was carried out in the posttest and in the follow-up. The order in which these versions were presented was counter-balanced, and participants were randomly assigned to one of the versions.

In the argument structure test, participants were asked to identify the different components of four short arguments. Before the actual test, another short argument was provided which served as an example. The example text did not include any information about the Toulmin model or explanations on how to deal with arguments, but was merely an additional text so participants could get familiar with the task. The participants were asked to read the complete text first. In a second step, the text was presented again in fragments which consisted of several paragraphs, whereby each paragraph represented a different component of the argument, i.e. claim, datum, warrant, backing, and rebuttal. The paragraphs were numbered and participants were instructed to assign each number to its corresponding argument component that had to be selected from a list appearing at the bottom of the screen. For each argument component, a short definition was provided.

Training

One week after the pretest, participants returned to the lab for the training intervention. As in the pretest, they were welcomed, briefly informed about the procedure upon arrival, and seated in front of a computer. Subsequently, they were randomly assigned to either the argument structure training intervention or to a control group in which they worked on their speed-reading competences.

Argument structure training

Participants in the argument structure training were allowed as much time as they needed to complete the training. They were provided with a headset that they were instructed to use during the video tutorial.

Participants received theoretical input first. After a short explanation of the relevance and purpose of arguments, the Toulmin (1958) model was introduced in a stepwise fashion using several examples, and the importance of markers and key words was highlighted. Subsequently, a number of learning goals were formulated, followed by two prompts in which participants were instructed to answer the questions formulated in the first and second learning goals. In the first exercise, they were asked to allocate each argument component to its corresponding position in the Toulmin (1958) model with the help of dropdown

elements. Immediate feedback was provided and participants were allowed to correct their responses, if necessary, or to proceed with the next task. In the second exercise, the argument components had to be entered in an empty text field. Again, participants received feedback on the correctness of their response and participants could either correct their response or continue. In the next step, participants were instructed to put their headsets on and watch the two video tutorials in which strategies to identify the components of arguments were demonstrated by a model. Following this, they were prompted to write down useful strategies to identify each argument component (third learning goal). They were allowed to access this information, along with the theoretical input and the tutorials, throughout the experiment by pressing a button at the bottom of each page (cp. Britt and Aglinskas 2002). In addition, this page appeared after every feedback, and participants could decide whether they wanted to review particular information or proceed. In the practical phase, a number of different arguments were presented. These arguments were preceded by an example text. Participants were instructed to select the appropriate argument component for each paragraph of a text that was presented as a complete text first, and then in fragments. In addition, they were asked to find markers and write them down in an empty text field. A scheme displaying the Toulmin (1958) model appeared at the bottom of each practice text. As soon as each argument component was assigned a position in a text, participants received feedback on the correctness of their response, and the correct solution appeared both in the text and in the scheme. Again, they were given the opportunity to correct their responses, review certain information (e.g., theoretical input, video tutorials, notes), or continue with the following text. Finally, participants were once again prompted to provide an answer to the three learning goals that had been formulated at the beginning of the experiment, and to write down which parts of the training they found most helpful, before they were allowed a short break (15 min).

Speed-reading training

Participants in the control group were provided with tablets and worked on eight exercises, whereby each of them was limited to a processing time of 60 s. The exercises included an initial assessment of reading speed (1), tasks whereby a moving dot had to be tracked while different words were presented (2), particular letters had to be identified (3, 6), dissimilar word pairs identified (4, 7), particular words tracked while fixating a row (5), and a task wherein a dot had to be followed along several rows of words. After the completion of all exercises, participants were shown how many points they had collected in each exercise and took a break of 25 min. After that, another eight, participants were instructed to keep practising with similar exercises until the timer reached 50 min. Finally, participants were allowed another 15-min break.

Posttest

After the break following the training intervention, participants completed the argument structure test again. They were randomly assigned to one of the parallel versions that they had not completed yet. At the end of the session, they were asked to indicate how confident they felt in dealing with the argument structure model on a Likert scale ranging from *1 = not confident at all* to *6 = very confident*. Finally, the students were thanked again for their participation and reminded of the upcoming follow-up session, before they were dismissed.

Follow-up

Both tests were completed a third time in the final session, whereby participants worked on the remaining parallel versions of the tests. They were once again asked about their confidence with regard to the argument structure model and its application at the end of the session. Finally, they were thanked for participation, reimbursed with course credits or financial remuneration, and dismissed. Participants were debriefed a few weeks later, and received an individual feedback about their training success upon request.

Design

The study comprised of a single factor (intervention: argumentation structure training versus speed-reading training) between-subjects design. Accuracy of responses in the posttest served as the dependent variable. Participants were randomly assigned to one training condition. The test battery included three parallel versions of the test with four short argumentative texts for the argument structure test in each version. The order in which the versions were presented was counter-balanced across participants. Differences in pretest accuracies and study performance were controlled for as covariates.

Data analysis strategy

Type-I-error probability was set at .05 for all hypothesis tests. One-tailed tests were used for directional predictions. Hypothesis 2 and Research Questions 1 and 2 were tested with posttest or follow-up accuracies (controlling for pretest accuracies) as the outcome variable.

For testing Hypothesis 2 and examining Research Questions 1 and 2, we used linear models with categorical and continuous predictors and interaction terms (Cohen et al. 2003, Chap. 9). All continuous predictors were z-standardised. Training condition was included as contrast-coded predictor (1: argument structure training, $-$ 1: speed-reading control condition). A sequence of two nested models were tested. In Model 1, training condition, pretest accuracies and their interactions were included as predictors. In Model 2, study performance and its interactions with training condition were added to the model.

For testing Hypothesis 3, paired-samples t-tests were conducted to examine whether confidence in the experimental group improved after the intervention, compared to the pretest.

Finally, Research Question 3 was examined by looking at frequencies of answers concerning the helpfulness of different elements of the constructivist learning environment.

Results

Pretest accuracy scores

Both training groups achieved similar accuracy scores in the pretest, $p > .05$. Atypical, full-fledged arguments ($M = .52$, $SE = .04$) were more difficult to identify than typical, full-fledged arguments ($M = .68$, $SE = .05$), $p < .001$, which, again, were more challenging than arguments with only three components ($M = .89$, $SE = .03$), $p < .001$. Thus, as predicted by

Hypothesis 1, complex arguments with a less typical structure were more challenging to identify than arguments with more typical structure.

Learning gain in the posttest and follow-up

The results for Model 1 (Table 1) showed that training in argument structure did not generally improve performance in the posttest or follow-up, $p > .05$. Furthermore, no significant differences between the training group and the control group were found at the posttest for the ability to identify claims ($B = 0.04$, $SE = 0.03$, $p > .05$, one-tailed), reasons ($B = 0.01$, $SE = 0.04$, $p > .05$, one-tailed), or rebuttals ($B = 0.00$, $SE = 0.02$, $p > .05$, one-tailed). However, in a more detailed analysis by argument components, a significant effect of training condition was revealed for the ability to identify warrants ($B = 0.11$, $SE = 0.04$, $p < .001$, one-tailed, $\Delta R^2 = .09$), with significantly improved accuracy values in the argument structure training group ($M = .64$, $SE = .06$), as compared to the speed-reading training group ($M = .41$, $SE = .06$). Furthermore, in the posttest, Model 2 revealed a main effect of training condition for the identification of atypical, full-fledged arguments ($B = 0.08$, $SE = 0.03$, $p < .01$, $\Delta R^2 = .09$, one-tailed), with participants in the experimental condition receiving higher posttest accuracies ($M = .68$, $SE = .04$), as compared to those in the control condition ($M = .53$, $SE = .04$). Thus, in partial support of Hypothesis 2, participants in the experimental group were able to improve their ability to identify less typical components and more complex arguments with a less typical structure. Unexpectedly, however, no such effects were found in the follow-up, $p > .05$.

Table 1 Summary of nested multiple regression analyses for variables predicting posttest performance after the training intervention

Variable	B	SE_B	t (df)	F (df_h, df_e)	R^2	ΔR^2
Step 1				3.45 (3,45)*	.19	
Intercept	.71	.02	29.56*** (46)			
Training condition (TC)	.04	.02	1.50 (46)			.04
Pretest accuracy (PA)	.06	.03	2.50** (46)			.10
PA* TC	.03	.03	1.01 (46)			.03
Step 2				6.09 (5,43)***	.41	
Intercept	.70	.02	33.93*** (44)			
Training condition (TC)	.04	.02	1.65 (44)			.04
Pretest accuracy (PA)	.04	.02	1.88*(44)			.05
PA* TC	− .01	.02	.39 (44)			.00
Study performance (SP)	− .08	.02	− 3.45** (44)			.17
SP* TC	− .05	.02	2.08* (44)			.06

$N = 50$. Pretest accuracy and study performance were z-standardized. Training condition was included as contrast-coded predictor (1: argument structure training, −1: speed-reading control condition). Note that lower values represent better performance in the German grading system

*$p < .05$. **$p < .01$ ***$p < .001$. One-tailed tests were used to test for effects of training condition and study performance, and for the interaction of study performance and training condition

Individual differences in pretest accuracies and study performance

In Model 1, a significant effect of pretest accuracies was found for the posttest ($B=0.06$, $SE=0.03$, $p<.01$, $\Delta R^2=.10$) and for the follow-up ($B=0.08$, $SE=0.02$, $p<.001$, $\Delta R^2=.23$), indicating that students with higher pretest accuracies scored higher after the intervention and 4 weeks later. When study performance and its interaction with training condition were added to the model (Model 2), study performance moderated the effect of training condition in the posttest, ($B=-0.05$, $SE=0.02$, $p<.05$, $\Delta R^2=.06$). To interpret the interaction, we estimated and plotted the simple slopes of study performance in the argument structure training and the speed-reading training condition (Fig. 1) and estimated the effect of training condition at a low level of study performance and at a high level of study performance (Cohen et al. 2003, Chap. 9). The negative slope of study performance was steeper in the argument structure training condition ($B=-0.12$, $SE=0.03$, $p<.001$, one-tailed, $\Delta R^2=.06$) than in the speed-reading condition where it was not significant ($B=-0.03$, $SE=0.03$, $p=.17$, one-tailed). Note that the simple slopes are negative, because lower values represent better performance in the German grading system. At a low level of study performance (i.e., a mean grade of 1 SD above the sample mean), the two training conditions did not differ in post-test accuracy (argument structure training: $M=.61$, $SE=.05$; speed-reading training: $M=.64$, $SE=.05$), $t(43)=-0.50$, $p=.31$, one-tailed. At a mean level of study performance, the post-test scores were higher after the argument structure training ($M=.73$, $SE=.03$) compared to the speed-reading training ($M=.67$, $SE=.03$), but the effect missed the significance criterion by a narrow margin, $t(43)=1.65$, $p=.05$, one-tailed. In contrast, at a high level of study performance (i.e., a mean grade of 1 SD below the sample mean), participants in the argument structure training clearly outperformed those in the speed-reading training (argument structure training: $M=.86$, $SE=.04$; speed-reading training: $M=.70$, $SE=.05$), $t(43)=2.43$, $p<.01$, one-tailed. Thus, it can be concluded with regard to the exploratory Research Question 2 that students with very good grade average, i.e. above-average study performance, benefitted from the argument structure training, although this effect could again not be shown in the follow-up, $p>.05$.
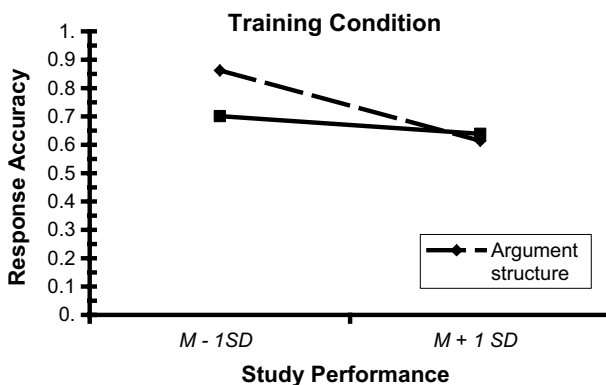

Fig. 1 Estimates of the simple slopes (with standard errors) of the effect of study performance on posttest accuracies, after participating in the argument structure training or the control training (speed-reading training)

Whereas the initial variables could only explain a moderate amount of the variance in our model ($R^2 = .19$), more than 40% of the variance could be explained after the addition of study performance and its interaction with training condition ($R^2 = .41$). In addition, as in Model 1, a significant main effect of pretest accuracies was found in the posttest ($B = 0.04$, $SE = 0.02$, $p < .05$, $\Delta R^2 = .05$) and at follow-up ($B = 0.07$, $SE = 0.02$, $p < .01$, $\Delta R^2 = .15$).

Results of the moderated regression analyses for the posttest are displayed in Table 1. To better understand these global effects that were found in the posttest, a number of follow-up analyses were performed, whereby Model 2 served as the basis for analysis.

Accuracy in different argument types

For typical, full-fledged arguments, pretest scores moderated the effect of training condition ($B = 0.09$, $SE = 0.04$, $p < .05$, $\Delta R^2 = .06$). Estimation of the simple slopes of pretest scores for each training condition showed that the negative slope of pretest scores was steeper in the argument structure training condition ($B = 0.20$, $SE = 0.06$, $p < 01$, one-tailed, $\Delta R^2 = .06$) compared to the speed-reading condition, where it was not significant ($B = 0.04$, $SE = 0.06$, $p = .27$, one-tailed). At a low level of pretest scores (i.e., a mean grade of 1 SD above the sample mean), the two training conditions did not differ in post-test accuracy (argument structure training: $M = .88$, $SE = .09$; speed-reading training: $M = .79$, $SE = .09$), $t (43) = 0.77$, $p = .22$, one-tailed. Again, no significant differences between the argument structure training ($M = .74$, $SE = .06$) and the speed-reading training ($M = .68$, $SE = .06$) were found at a mean level of pretest scores, $t (43) = 0.17$, $p = .25$, one-tailed. In contrast, at a high level of pretest scores (i.e., a mean grade of 1 SD below the sample mean), participants in the argument structure training performed better than those in the speed-reading training (argument structure training: $M = .69$, $SE = .09$; speed-reading training: $M = .48$, $SE = .09$), $t (43) = 1.75$, $p < .05$, one-tailed. Thus, students with high pretest scores particularly benefitted from the argument structure training with regard to the ability to identify typical, full-fledged arguments. No significant training effects were observed for arguments with three components ($B = .01$, $SE = .03$, $p > .05$). These results provide further clarification for Research Question 1.

Accuracy in different argument components

Study performance moderated the effect of training condition for identifying backing evidence ($B = -0.09$, $SE = 0.05$, $p < .05$, $\Delta R^2 = .06$). Estimation of the simple slopes of study performance for each training condition showed that the negative slope of study performance was steeper in the argument structure training condition ($B = -0.16$, $SE = 0.06$, $p < .01$, one-tailed, $\Delta R^2 = .06$) compared to the speed-reading condition, where it was not significant ($B = .01$, $SE = .06$, $p = .44$, one-tailed). At a low level of study performance (i.e., a mean grade of 1 SD above the sample mean), the two training conditions did not differ in posttest accuracy (argument structure training: $M = .46$, $SE = .09$; speed-reading training: $M = .40$, $SE = .09$), $t (43) = -0.38$, $p = .34$, one-tailed. Similarly, posttest accuracies did not differ significantly at a mean level of study performance between the argument structure training ($M = .54$, $SE = .06$) and the speed-reading training ($M = .45$, $SE = .06$), $t (43) = -0.95$, $p = .17$, one-tailed. However, at a high level of study performance (i.e., a

mean grade of 1 *SD* below the sample mean), participants in the argument structure training outperformed those in the speed-reading training (argument structure training: $M = .67$, $SE = .09$; speed-reading training: $M = .45$, $SE = .09$), $t (43) = -1.69$, $p < .05$, one-tailed. Thus, students with very good average grades, i.e. above-average study performance, particularly benefitted from the argument structure training with regard to their ability to identify backing evidence. These results further inform Research Question 2.

Self-reported confidence with the Toulmin model

As predicted in Hypothesis 3, participants in the experimental group felt significantly more confident with the Toulmin model after the intervention ($M = 4.75$, $SE = .14$), compared to the pretest ($M = 3.17$, $SE = .16$), $t (23) = -10.00$, $p < .001$, one-tailed. Moreover, although their confidence dropped a little at follow-up, they still felt more confident than in the pretest ($M = 3.67$, $SE = .18$), $t (23) = -3.39$, $p < .01$, one-tailed.

Helpfulness of CLE elements

When asked which parts of the training experiment participants found most helpful for improving their competence to recognize different argument components, the video tutorials were named most often (15), followed by the practice phase and the feedback (both 9), the theoretical input (6), and the prompts (1). Thus, regarding the exploratory Research Question 3, different elements from the constructivist learning environment were perceived as helpful by the participants in the experimental group, but the video tutorials were perceived as most helpful, followed by the practical exercises and the feedback.

Discussion

The present experiment investigated how training in the ability to recognise different structural components of arguments could improve psychology students' competences to comprehend informal arguments. Results indicate that familiarising students with the structure of arguments improved their ability to recognise warrants and more complex (full-fledged) or less typical arguments (Toulmin 1958). Generally, students felt more competent with the Toulmin model after the intervention. Different elements from the constructivist learning environment, such as video-based tutorials, the practice phase, and the presence of feedback, were perceived as helpful. Students with very good grades particularly profited from the training intervention, as reflected in significantly improved performances after the intervention for those who participated in the argument structure training. Moreover, students who were initially able to recognise more complex argument types could further improve this ability in the intervention. Our results suggest that shifting attention towards relational aspects between argument components (i.e. warrants) showed the greatest increment in students' posttest performance. Thus, acquisition of conceptual and procedural knowledge about informal arguments may have helped with the formation of accurate representations of key components of arguments, including warrants.

The students who participated in the argument structure training were generally able to improve their performance to recognise less typical argument components, such as warrants, and more complex (full-fledged) arguments with a less typical structure. However,

participants in both groups were already relatively accurate in their ability to recognise more typical components, such as rebuttals (89% accuracy), and, to a lesser degree, claims (62% accuracy), and data (62% accuracy) or less complex argument types, such as arguments with only three components (89% accuracy), prior to the intervention. These results indicate that the students seemed to possess some prior knowledge of the structure of (less complex) arguments, and that our results were likely affected by ceiling effects. However, only a minority of the participants in our study were able to correctly identify warrants (35% accuracy). Accuracy values for the identification of warrants almost doubled after the intervention for those who participated in the argument structure training (64%), suggesting that the intervention especially improved awareness of relational aspects between argument components. Thus, training may be especially useful for less typical components, such as warrants, and for more complex, full-fledged arguments with a less typical structure (e.g. reason-first arguments, Britt and Larson 2003). Our results are in line with previous research indicating that students tend to neglect the internal consistency of arguments (e.g., Britt and Kurby 2005; Larson et al. 2004, 2009; Shaw 1996; von der Mühlen et al. 2016), but that training in argument structure can be effective in overcoming these deficits (e.g., Larson et al. 2004).

The results found in our study can be interpreted in the framework of the mental model theory (Johnson-Laird 1983). Whereas arguments with more typical components and a typical claim-first structure are likely to be congruent with the current state of the reader's mental model (Schroeder et al. 2008), arguments with less typical components and a less typical structure seem to be more challenging to deal with. We presume that training in the identification of structural components of arguments, including components signaling relations between key components (i.e. warrants), allowed the construction of more accurate representations of arguments in memory and helped students to activate different argument components simultaneously when trying to understand these arguments (Britt et al. 2014; Shaw 1996).

Not everyone profited from the training intervention to the same extent. Students with a better study performance profited the most from training in argumentation. We assume that students who performed very well in their current education were more familiar with a broad range of scientific texts than the average student, which might have (implicitly) provided them with some relevant background knowledge (i.e. *discipline expertise*, Rouet et al. 1997) about the structure of arguments. This knowledge might have allowed them to more easily comprehend, integrate and apply information from the training intervention. In addition, their experience with different scientific literature might have facilitated their argument comprehension skills (Britt et al. 2014; Rouet et al. 1997). In particular, students with very high study performance especially improved their competence to identify backing evidence for warrants, indicating that these students paid particular attention to less typical, relational aspects of arguments.

In our experiment, we also examined effects of pretest accuracies on posttest performance. Students with already high initial accuracy scores scored significantly higher after the intervention, indicating that they could further improve their ability to identify key components of arguments, especially warrants, and more complex, full-fledged arguments. Thus, students who were already able to recognise more complex types of arguments could further improve this competence during the training intervention. These students, similarly to those with high study achievement, were likely to possess some relevant prior knowledge about arguments. This knowledge might have helped them to concentrate on acquiring further knowledge about less typical argument components and more complex arguments. Our analyses showed that both study performance and pretest accuracies

contributed independently to the variance in our model, suggesting that they might involve different kinds of prior knowledge, or that one or the other might reflect a more general cognitive ability, such as individual differences in the ability to think rationally (Stanovich 2012).

It is not fully understood why the students with very good study performance could profit most from our intervention. While we assume that these students have acquired more experience with the structure of scientific texts and arguments, other mechanisms, such as differences in intelligence, might be responsible for the observed interaction of training condition and study performance. Although Stanovich (2012) found that the skill of rational thinking seems to be independent of intelligence, future research should address this issue.

In addition, although the majority of students in our sample were undergraduate students in their second semester, the sample also included some more advanced students. The possibility cannot be ruled out that our results were to some degree influenced by differences in age or experience with scientific texts and arguments.

Furthermore, the present study does not flesh out the precise mechanisms under which students acquire a reasoning schema. Although students indicated that they perceived the video tutorials, the practice phase, and the presence of feedback as very helpful, manipulations that include or exclude different tools and measures tapping into cognitive processes during training would be necessary to achieve insights that are more objective.

It should also be noted that the results observed in our intervention could not be replicated 4 weeks later, indicating that a single session may not be enough to produce long-term effects. Expertise takes time to develop (Britt et al. 2014; Ericsson et al. 1993). Therefore, future research should examine more extensive interventions that include several practice sessions and integrate such interventions into the curriculum.

Moreover, future studies should investigate whether our training is also effective for students with other levels of experience, such as younger and less advanced students (e.g., high school and college students), or with more advanced students (e.g., graduate students). Such studies should also match learners' abilities to enable successful learning and prevent ceiling effects. Ideally, they should also include a larger sample.

Despite these limitations, our results indicate that interventions focusing on the construction of conceptual and procedural knowledge about informal arguments in a constructivist setting can be effective for fostering students' competences to comprehend these arguments. Understanding relations between key components of arguments are important prerequisites for the successful evaluation of the quality of arguments (Blair and Johnson 1987; Voss and Means 1991).

Finally, our results raise the question of how much value we assign to the acquisition of epistemic competences in formal instruction and education. Assuming that lack of practice is one of the main reasons why students find it difficult to comprehend (more complex) arguments (Perkins 1985; Perkins et al. 1991), interventions to foster argumentation skills should be included into the curriculum to help students develop argument schemes that become activated when needed to guide comprehension. Such interventions should also be designed to match the characteristics of learners (Snow 1989). Requiring students to read various scientific documents on a regular basis may be a first step to allow the construction of some relevant structural knowledge, which, in turn, could help them to particularly profit from further training in the skill of argumentation.

Appendix: Sample Item for a Typical Argument (Argument 1) and a Less Typical Argument (Argument 2) Used in the Video Tutorial (Translated from German)

Argument 1

We can help children who suffer from nightmares with a simple method [*claim*]. In this regard, results from a study showed that repeated drawings of the threatening content in the nightmare (e.g., Dracula) and tearing them apart afterwards made the nightmare disappear [*datum*]. The procedure is simple, because it can be easily used by parents [*warrant*]. For example, parents can integrate it as part of their daily night-time routines [*backing*]. For the method to be successful, however, it is important that the child is ready to confront herself with her fears [*rebuttal*].

Argument 2

Results from a study indicate that women showing a confident, dominant appearance during the application procedure receive lower ratings for their social competence [*datum*]. Social competence is in great demand on the job market [*warrant*]. In a different study, it was found that social competence was an important criterion in 70% of application processes [*backing*]. Thus, high expectations for social competence may ironically create more discrimination in women [*claim*], provided they present themselves as confident career women [*rebuttal*].

References

Amstad, T. (1978). *Wie verständlich sind unsere Zeitungen? [How understandable are our newspapers?]* (Unpublished doctoral dissertation). Zürich: University of Zürich.

Beck, I. (1989). *Reading today and tomorrow*. Austin, Texas: Holt and Co.

Berthold, K., & Renkl, A. (2010). How to foster active processing of explanations in instructional communication. *Educational Psychology Review, 22,* 25–40.

Blair, J. A., & Johnson, R. H. (1987). Argumentation as dialectical. *Argumentation, 1,* 41–56.

Britt, M. A., & Aglinskas, C. (2002). Improving student's ability to use source information. *Cognition and Instruction*, *20*, 485–522.

Britt, M. A., & Kurby, C. A. (2005). *Detecting incoherent informal arguments*. Paper presented at the 15th annual meeting of the Society for Text and Discourse, Chicago, IL.

Britt, M. A., Kurby, C. A., Dandotkar, S., & Wolfe, C. R. (2008). I agreed with what? Memory for simple argument claims. *Discourse Processes, 45,* 52–84.

Britt, M. A., & Larson, A. (2003). Construction of argument representations during on-line reading. *Journal of Memory and Language, 48,* 749–810.

Britt, M. A., Richter, T., & Rouet, J.-F. (2014). Scientific literacy: The role of goal-directed reading and evaluation in understanding scientific information. *Educational Psychologist, 49,* 104–122.

Calfee, R., & Chambliss, M. (1988). Beyond decoding: Pictures of expository prose. *Annals of Dyslexia, 38,* 243–257.

Chambliss, M. J. (1995). Text cues and strategies successful readers use to construct the gist of lengthy written arguments. *Reading Research Quarterly, 30,* 778–807.

Chambliss, M. J., & Murphy, P. K. (2002). Fourth and fifth graders representing the argument structure in written texts. *Discourse Processes, 34,* 91–115.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13,* 145–182.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.

De Winstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory and Cognition, 32,* 945–955.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology, 18,* 105–115.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100,* 363–406.

Freeman, A. M., Morris, P. B., Barnard, N., Esselstyn, C. B., Ros, E., Agatston, A., et al. (2017). Trending cardiovascular nutrition controversies. *Journal of the American College of Cardiology, 69,* 1127–1187.

Galotti, K. M. (1989). Approaches to study formal and everyday reasoning. *Psychological Bulletin, 105,* 331–351.

Hefter, M. H., Berthold, K., Renkl, A., Rieß, W., Schmid, S., & Fries, S. (2014). Effects of a training intervention to foster argumentation skills while processing conflicting scientific positions. *Instructional Science, 42*(6), 929–947.

Hefter, M. H., Renkl, A., Riess, W., Schmid, S., Fries, S., & Berthold, K. (2015). Effects of a training intervention to foster precursors of evaluativist epistemological understanding and intellectual values. *Learning and Instruction, 39,* 11–22.

Heku-IT [computer software]. http://www.heku-it.com/schneller-lesen-app/.

Inquisit (version 3.0.6.0) [computer software]. http://www.millisecond.com/.

Johnson, B. T., Smith-McLallen, A., Killeya, L. A., & Levin, K. D. (2004). Truth or consequences: Overcoming resistance with positive thinking. In E. S. Knowles & J. A. Linn (Eds.), *Resistance and persuasion* (pp. 215–233). Mahwah, NJ: Erlbaum.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness*. Cambridge, MA: Harvard University Press.

Jonassen, D. (1999). Designing constructivist learning environments. In C. M. Reigeluth (Ed.), *Instructional design theories and models* (pp. 215–239). Hillsdale, New Jersey: Lawrence Erlbaum.

Jonassen, D. H. (1993). Cognitive flexibility theory and its implications for designing CBI. In S. Dijkstra, H. P. Krarnmer, & J. V. Merrienboer (Eds.), *Instructional models in computer based learning environments*. Heidelberg: Springer.

Kilpatrick, J., Swafford, J. O., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Larson, A. A., Britt, M. A., & Kurby, C. (2009). Improving students' evaluation of informal arguments. *Journal of Experimental Education, 77,* 339–365.

Larson, M., Britt, M. A., & Larson, A. (2004). Disfluencies in comprehending argumentative texts. *Reading Psychology, 25,* 205–224.

Marsh, E. J., Edelman, G., & Bower, G. H. (2001). Demonstrations of a generation effect in context memory. *Memory and Cognition, 29,* 798–805.

Mayer, R. (1989). Models for understanding. *Review of Educational Research, 59,* 43–64.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87,* 319–334.

National Assessment of Educational Progress. (1996). NAEP 1994 U.S. History report card: Findings from the national assessment of educational progress. Princeton, NJ: Educational Testing Service.

OECD. (2011). *PISA 2009 results: Students on line: Digital technologies and performance* (Vol. 6). Paris: OECD. http://dx.doi.org/10.1787/9789264112995-en.

OECD. (2014). *PISA 2012 results: What students know and can do—Student performance in mathematics, reading and science* (Vol. 1, Revised ed.). Paris: PISA, OECD Publishing. http://dx.doi.org/10.1787/9789264208780-en.

Paxton, R. J. (1997). "Someone with like a life wrote it": The effects of a visible author on high school history students. *Journal of Educational Psychology, 89,* 235–250.

Perkins, D. N. (1985). Postprimary education has little impact on informal reasoning. *Journal of Educational Psychology, 77,* 562–571.

Perkins, D. N., Farady, M., & Bushey, B. (1991). Everyday reasoning and the roots of intelligence. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), *Informal reasoning and education* (pp. 83–105). Hillsdale, NJ: Erlbaum.

Renkl, A. (2009). Wissenserwerb [Knowledge acquisition]. In E. Wild & J. Möller (Eds.), *Pädagogische Psychologie* (pp. 3–26). Berlin: Springer.

Richter, T. (2011). Cognitive flexibility and epistemic validation in learning from multiple texts. In J. Elen, E. Stahl, R. Bromme, & G. Clarebout (Eds.), *Links between beliefs and cognitive flexibility*. Berlin: Springer.

Rittle-Johnson, B., Star, J. R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology, 101,* 836–852.

Rouet, J.-F., Favart, M., Britt, M. A., & Perfetti, C. A. (1997). Studying and using multiple documents in history: Effects of discipline expertise. *Cognition and Instruction, 15,* 85–106.

Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching, 41,* 513–536.

Schroeder, S., Richter, T., & Hoever, I. (2008). Getting a picture that is both accurate and stable: Situation models and epistemic validation. *Journal of Memory and Language, 59,* 237–255.

Shaw, F. W. (1996). The cognitive processes in informal reasoning. *Thinking and Reasoning, 2,* 51–80.

Snow, R. E. (1989). Aptitude-treatment interaction as a framework of research in individual differences in learning. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences* (pp. 13–59). New York: Freeman.

Stanovich, K. E. (2012). On the distinction between rationality and intelligence: Implications for understanding individual differences in reasoning. In K. Holyoak & R. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 343–365). New York: Oxford University Press.

Stiftung Warentest. (2015). Lesetrainings im Test: Wie Sie zum Schnellleser werden [Reading trainings tested: How to become a fast reader]. https://www.test.de/Lesetrainings-im-Test-Wie-Sie-zum-Schnellleser-werden-4817442-0/.

Suppe, F. (1998). The structure of a scientific paper. *Philosophy of Science*, *65*(3), 381–405.

Toulmin, S. E. (1958). *The uses of argument*. Cambridge, MA: Cambridge University Press.

von der Mühlen, S., Richter, T., Schmid, S., Schmidt, E. M., & Berthold, K. (2016). Judging the plausibility of argumentative statements in scientific texts: An student-scientist comparison. *Thinking and Reasoning, 22,* 221–246.

Voss, D., Perkins, N., & Segal, J. W. (Eds.). (1991). *Informal reasoning and education* (pp. 83–106). Hillsdale, NJ: Erlbaum.

Voss, J. F., & Means, M. L. (1991). Learning to reason via instruction in argumentation. *Learning and Instruction, 1,* 337–350.

Wolfe, C. R., Britt, M. A., & Butler, J. A. (2009). Argumentation schema and the Myside Bias in written argumentation. *Written Communication, 26,* 183–209.