Developing Pre-Service Teachers' Evidence-Based Argumentation skills on Socio-Scientific Issues

Kalypso Iordanou[a] and Costas P. Constantinou[b]

[a]University of Central Lancashire Cyprus

12-14 University Avenue

Pyla, 7080 Larnaka, Cyprus

Tel: +357 24 81 21 21 Fax: +357 24 81 21 20

Email: KIordanou@uclan.ac.uk

[b]University of Cyprus, Cyprus

Learning in Science Group

Department of Educational Sciences

P. O. Box 20537, CY 1678

Tel. +35722892936, Fax +35722333778

Email: c.p.constantinou@ucy.ac.cy

Corresponding author: Kalypso Iordanou

Email: KIordanou@uclan.uk.ac

Abstract

We report on a study of the effect of meta-level awareness on the use of evidence in discourse. The participants were 66 pre-service teachers who were engaged in a dialogic activity. Meta-level awareness regarding the use of evidence in discourse was heightened by having same-side peers collaborating in arguing on the computer against successive pairs of peers on the opposing side of an issue on the topic of Climate Change and by engaging in explicit reflective activities on the use of evidence. Participants showed significant advances both in their skill of producing evidence-based arguments and counterarguments and regarding the accuracy of the evidence used. Advances were also observed at the meta-level, reflecting at least implicit understanding that using evidence is an important goal of argumentation. Another group of pre-service teachers, who studied about the role of evidence in science in the context of regular curriculum and served as a control condition, did not exhibit comparable advances in the use of evidence in argumentation. Educational implications are discussed.

Keywords: Argumentation, reasoning, evidence, metacognition, science.

Developing Pre-Service Teachers' Evidence-Based Argumentation skills on Socio-Scientific Issues

1. Introduction

Argument skills have a central role in science education (Kuhn, 2010; Lehrer, Schauble, & Petrosino, 2001). Advanced argument skills are associated with appreciation of science as an enterprise that advances through coordination of evidence with theories, rather than as the accumulation of facts (Sandoval, 2005). The goal of advanced argument skills, according to Walton (1989), is to weaken the opponent's claims. Key means to achieving this goal are the employment of argumentive strategies that critique the opposing claim and the effective use of evidence (Kuhn, Zillmer, Crowell & Zavala, 2013). Yet, studies of students' argument skills in science contexts report these skills to be under-developed at best, even at the college level (Kelly, Druker, & Chen, 1998; Kelly & Takao, 2002; Kolstø, Bungum, Arnesen, et al., 2006; Maloney & Simon, 2006; Sadler, 2004; Maloney, 2007). Despite this fact, the available empirical research in science education studying how the development of argument skills can be supported, has been limited. Until recently, most research on scientific reasoning has been devoted to control of variables, and much less attention has been paid to argument skills (Kuhn, Iordanou, Pease & Wirkala, 2008). Key questions such as the degree to which educational interventions can promote the development of students' argument skills, and especially their ability to use evidence in argumentation, remain unresolved.

Evidence lies at the heart of science. It constitutes the foundation of science and the mechanism through which it advances. A theory's power and its potential influence in the field of physics are determined by reasoned evaluation of the available data (Franklin, 1994). Evidence evaluation lies not only at the heart of science, but it has a prominent position in scientific reasoning in the personal arena. Students' approach towards evidence is identified as a central element of epistemological understanding, that is an understanding of the constructive nature of knowledge (Kuhn et al., 2008), which relates to self-regulated learning (Muis, 2007). The kind of evidence that students value as good evidence in a particular context, what Chinn, Buckland and Samarapungavan (2011) refer to as the evidential standards students pursue, influence the learning processes that students engage in and the conclusions they draw. Yet, studies repeatedly show that students do not use evidence consistently to support their claims (Erduran et al., 2004; Jiménez -Aleixandre, Rodriguez, & Duschl, 2000), even when they are explicitly instructed to do so (Sandoval & Millwood, 2008). Whether this phenomenon could be due to lack of skill, lack of appreciation of the need to do so or both, remains an open issue.

Instead of coordinating theory with evidence, including changing prior theory where necessary in the light of new evidence, students tend to rely excessively on their initial theories in their justifications. Even in cases where their prior beliefs are inconsistent with the available data, students still tend to make judgments that are consistent with their prior beliefs (Amsel & Brock, 1996). This over-reliance on prior beliefs is well documented in multiple studies across the fields of social psychology (Petty & Wegener, 1999), rational thinking – known as the myside bias − (Stanovich, West, & Toplak, 2013) and science education (Schauble, 1990). In particular, in a study reported by Schauble (1990), elementary school students interpreted identical patterns of evidence differently, depending on whether a particular piece of evidence was consistent or inconsistent with their prior beliefs. In the latter case, students tend to misread, misinterpret or distort evidence in order to conclude that a particular piece of evidence supported their initial theories. In a more specific analysis of how people – both science students and scientists − respond to data that are inconsistent with one's own theory, Chinn and Brewer

(1993) identified seven possible reactions. These are ignoring, rejecting, excluding anomalous data from the current theory, holding anomalous data in abeyance, reinterpreting them, making peripheral theory change, and finally changing one's theory. Note that only one, out of the seven responses listed, involves theory change to accommodate new data.

The way students handle data does not seem to change with age. Even though young children show some competence in producing arguments in support of a claim (Anderson, Chinn, Chang, Waggoner, & Yi, 1997; Stein & Miller, 1993), serious weaknesses have been observed in the arguments of adolescents, adults, and even college students who major in Physics. Studying high school students' usage of justification for their claims as they were engaged in an electricity-based performance assessment, Kelly, Druker and Chen (1998) found that although students produced some warrants when involved in experimentation, when they formalized their reasons in a written form they didn't do so. In this study, Toulmin's (1958) term of "warrant" was employed, which is defined as a "statement that acts to show that the move from data to claim is valid" (p.856).

In addition, little improvement is observed in individual's ability to use evidence in argumentation with standard curriculum. Observations of high school students' discussions in science classrooms showed a predominance of claims and rare usage of evidence to back up claims (Jiménez-Alexander et al., 2000). Studies that have undertaken explicit teaching of argument skills in a scientific context have shown, that the ability of considering alternative positions and integrating evidence with claims, which are key components of argumentation, showed little improvement (Mercer & Littleton, 2007; Osborne, Erduran, & Simon, 2004; Zohar & Nemet, 2002). Of particular concern are findings revealing weaknesses in science teachers' understanding and performance in incorporating evidence in argumentation (Zohar, 2004). Zembal-Saul et al. (2002) reported science teachers' inability to determine what counts as evidence in the context of an investigation. These findings, along with the fact that our understanding of how science teachers' argument skills can be developed is very limited (Zohar, 2008), point to the urgent need for examining ways to support science teachers' argument skills. How can we expect teachers to support the development of students' argument skills if they haven't themselves developed adequately these skills? Or, why are we expecting teachers to promote evidence-based argumentation, if they do not see the point of it? In the present study, we examined the development of evidence-based argumentation skills in a sample of pre-service teachers. Although use of evidence is a key component of skilled argumentation (Kuhn et al., 2013), we use the term evidence-based argumentation to stress the focus of the present study in developing the skill of using evidence in argumentation. Evidence-based argumentation involves both the ability to use evidence in argument production, that is when one uses evidence to support his/her own position and the evaluation of evidence in argument reception, that is when one weighs evidence offered by the opponent. In particular, this study examined the effect of enhancing meta-level awareness of the use of evidence, along with engagement in peer discourse, on the use of evidence in argumentation. We first present empirical evidence regarding the development of argument skills, particularly the skill of using powerful argumentive strategies. Then, we present a theoretical model which proposes a relationship between meta-level knowledge and argument skills, followed by empirical evidence regarding the relationship between meta-level knowledge and the development of effective argumentive strategies. We end our literature review with a description of the present study and our hypothesis.

1.1. Development of Argument Skills

According to Walton (1989), skilled argumentation has two goals. One is to secure commitments from the opponent that can be used to support one's own argument. The other is to undermine the opponent's position by identifying and challenging weaknesses in his or her argument. Note that both goals require attention to the opponent's claims. Previous cross-sectional research showed that there were developmental differences between adolescents' and adults' dialogic argumentation skills in respect to Walton's goals of skilled argumentation. Adolescents and adults differed in their focus during argumentation and on the argumentation strategies that they employed. Adolescents focused on exposition of their own position and used relatively weak argumentive strategies, whereas adults focused on the other's position, trying to weaken it using powerful argumentive strategies (Felton & Kuhn, 2001).

A series of experimental studies, over the last fifteen years, have shown that dense engagement in dialogic argumentive discussions support the development of argument skills of adolescents towards the direction of adults' argument skills (Kuhn et al., 2013; Reznitskaya, Anderson, & Kuo, 2007; Crowell & Kuhn, 2014; Iordanou 2010, 2013; Felton, 2004; Felton & Kuhn, 2001; Kuhn, Goh, Iordanou & Shaenfield, 2008; Mason, 1998). Advancements have been observed both in individuals' written arguments, namely individual argument skills, and the strategies they employed when engaged in a dialogue with individuals holding opposing position, namely dialogic argumentation skills (Kuhn, 1991). In particular, Kuhn, Shaw and Felton (1997) showed the effectiveness of engagement in dyadic discussion on the development of argumentation skills. In their study, young adolescents and adults participated in five weekly dyadic discussions on a social topic with different classmates, while another group of participants received initial and final assessment at the same time interval as experimental condition but did not participate in dyadic discussion. Results showed that participants who engaged in dyadic discussions progressed from 1-sided to 2-sided arguments, showing a shift in their attention from one's own position to the other's position. In contrast, control condition participants did not show any progress in argumentation skills.

Felton and Kuhn (2001) examined whether arguing in the context of an agreeing dyad as opposed to a disagreeing dyad would influence the development of argumentation skills. Young adolescents and young adults were assigned to agreeing or disagreeing dyads and were asked to discuss their views on the topic of Capital Punishment over the course of five dialogues. The agreeing dyads were asked to find all the reasons they agreed on, while the disagreeing dyads were asked to resolve their differences of opinion and come to agreement if they could. Results showed that participants in disagreeing dialogues produced more powerful argumentive strategies, such as counter-arguments, than participants in the agreeing dialogues.

Kuhn and Udell (2003) examined the effect of engagement in dialogic argumentation with peers who held on opposing view on the topic in addition to engagement in an argument-focused activity with peers who held the same opinion. Another group of participants engaged only in an argument-focused activity (e.g., generating and evaluating reasons) and served as a control condition. In that study, experimental condition students collaborated with students who held the same opinion in developing arguments to support their position and then engaged in argumentive discourse with peers who held on opposing view on the topic. Results showed that the experimental condition was more effective in developing students' advanced argumentive strategies, such as counter-argument and rebuttal, than control condition. In addition, the Kuhn and Udell study (2003) highlighted the role of the goal-based component of their intervention in developing argument skills. Participants in the study who were instructed that they would participate in a final showdown, exhibited a greater frequency of usage of advanced argumentive strategies, counter-arguments, than participants who engaged in practice alone (Kuhn et al., 1997).

Overall, these studies show that dense practice in dialogic argumentation is an effective condition for supporting participants' ability to shift their attention from their own position to attending more to the opponent's position (Kuhn et al., 1997; Reznitskaya et al., 2007). Furthermore, engagement in goal-based dialogic argumentation activity with peers who hold an opposing view on a topic is even more effective in developing participants' powerful argumentive strategies than practice of argumentation alone or engagement in argumentation where opposition is missing (Kuhn & Udell, 2003; Felton & Kuhn, 2001).

1.2. Meta-level Awareness and Development of Argument Skills

To explain the development of argument skills, Kuhn (2001) has proposed a model in which development of strategies at the procedural level is supported by development at the meta-level. In particular, in the theoretical framework she proposed about intellectual performance, Kuhn identified two kinds of meta-level knowing which support the execution of the knowing strategies, such as argument and inquiry. The first meta-level knowing is procedural meta-knowing, which refers to meta-level understanding and management of the task and strategies available in an individual's repertoire. Procedural meta-knowing answers questions such as "What do knowing strategies accomplish?" and "When, where, why to use them?" The second meta-level knowing is declarative meta-knowing, representing one's epistemological understanding regarding knowledge and knowing. This declarative or epistemic meta-knowing determines whether knowing strategies are executed. Epistemic meta-knowing informs intellectual values - which deal with questions such as “Is there a point to arguing?” - and then values, in turn, determine which strategies are applied. Work such as Mason and Scirica’s (2006) has provided empirical evidence for the relation between declarative or epistemic meta-knowing and argumentation skills, but it is out of the scope of the present work to review this literature.

In the present study we focused on the relation between procedural meta-knowing, especially meta-strategic knowledge, and argumentation skills. A series of studies has shown that interventions aiming to enhance procedural meta-knowing, supported the development of powerful argumentive strategies and reasoning skills (Zohar & Pelet, 2008; Zohar & Ben David, 2008; Felton, 2004; Kuhn et al., 2013; Crowell & Kuhn, 2014; Kuhn et al.,, 2008; Iordanou, 2010; 2013). In particular, Felton (2004) showed that engagement in reflective activities that aimed to support meta-level awareness, along with discourse, is a more effective condition for fostering the development of powerful argumentive strategies than engagement in discourse alone. Participants who engaged in reflection and practice on the topic of capital punishment exhibited greater advancements in the use of powerful argumentive strategies than control group participants who engaged only in practice in argumentation on the same topic. These advancements were reflected in the increase observed in advanced argumentive strategies, which addressed the opponents’ position, and the decrease in the use of less effective argumentive strategies, which served to elaborate on one’s own position. Notably, only participants who engaged in both reflection and practice showed an increase in the use of counter-argument strategies on the transfer topic (abortion). According to Felton, “In order to transfer the use of counter-argument, participants had to understand the relative value of the strategies independently of the topic of argument. They had to understand its function in the context of argumentative discourse." (p. 45) Consistent with these findings, are the findings of Iordanou (2010) which showed that the gains in the use of argumentative strategies were not only evident in the context of the knowledge domain in which participants received practice and reflection, but transferred across knowledge domains (Iordanou, 2010) and communication modes (Iordanou, 2013). In the Iordanou (2010) study, forty sixth-graders engaged in electronic dialogues with peers on a controversial topic. Half the students debated a science topic (dinosaur extinction) while the other half debated a social topic (home schooling). Over 13 sessions, participants worked with a partner in arguing against different pairs of classmates who held an opposing view on the topic. In addition, they engaged in some reflective activities based on dialogue transcriptions. Another 18 sixth-graders were placed in a control (non-intervention) condition. The intervention proved effective in developing participants' argumentation skills in the domain in which it was carried out. Participants exhibited increased frequency of usage of advanced argument strategies and decreased frequency of less advanced - exposition - strategies within the context of their intervention topic. This method proved also to be successful in producing transfer of argument skills across domains in both directions - from a science to social topic and from a social to science topic, and thus establishing the generality of such skills across different content domains.

Kuhn et al. (2008) examined through the microgenetic method whether meta-level understanding about argumentive strategies develop and support the development of powerful strategies at the procedural level. In this study, meta-level awareness was supported by conducting argumentation on the computer, engaging participants in reflective activities on the use of argumentive strategies and having participants who held the same position on an issue collaborating in arguing against pairs who held the opposing position. Results showed that besides gains in argumentive strategies, gains were also observed in meta-level communications about the discourse, supporting Kuhn's (2001) model. Participants who exhibited meta-level understanding of the objectives of argument and specifically of the relevance of the other person's position employed strategies at the procedural level that directly addressed the other person's position.

Research on argumentation thus far, has provided important insights regarding our understanding of the development of the skill of employing powerful argumentive strategies, documenting that adolescents and adults typically concentrated their attention on exposition of their own claims, ignoring the opponent's position, but after engagement in a series of argumentive and reflective activities they paid more attention to the opponent's position, as it is reflected in the increased usage of counterarguments and rebuttals. Yet, prior research examining the effect of multifaceted interventions on participants' argumentation skills (Iordanou, 2010; Kuhn & Udell, 2003; Kuhn et al., 2008) has focused on examining participants' attention in argumentation, without examining students' employment of evidence to back up their arguments and counterarguments. An exception of this line of work is the work of Kuhn et al., (2013) that examined the effect of extended engagement in argumentation on evidence use in construction of individual arguments. In that study, participants engaged in peer argumentation twice a week over the course of three years, having available relevant evidence. Kuhn and her colleague's findings of some improvement in adolescents' use of evidence in individual argument, showing improvements in using evidence to support a claim but not in arguing against a claim, are promising but at the same time point to the challenge of developing evidence-based argumentation. Therefore the question of whether enhanced meta-level awareness of evidence use in argumentation could support individuals' skill in employing evidence in dialogic argumentation to back up their own claims and the critique they offer to the opponent's claims, still remains unanswered.

1.3.The Present Study

In the present work, we extend previous research by examining the development of evidence-based argumentation in a group of undergraduate students (pre-service teachers). In particular, we examined the effect of enhancing meta-level awareness of the use of evidence in discourse, along with engagement in peer discourse, on the use of evidence in argumentation.

According to Kuhn's (2001) model, meta-level awareness is a necessary condition for formulating, and ultimately implementing goals with respect to it. Previous work which supported meta-level awareness of the goals of argumentive strategies showed that this awareness had a positive effect on the development of powerful strategic moves that constituted the dialogue (Kuhn et al., 2008). In particular, an increase in frequency was observed of counter-arguments, which critiqued the other's arguments, and rebuttals, which critiqued the opponent's counter-arguments to one's own claims. In the present work, we examined whether enhancing meta-level awareness of the use of evidence in particular, would support the development of evidence-based argumentation skills.

To heighten and support meta-level awareness of the use of evidence we employed explicit reflective activities which prompted students to reflect on (a) whether they had used evidence in the arguments and counter-arguments they had produced, (b) the quality of the evidence used and also think of how they could improve their evidence-based arguments and counter-arguments (see sections 2.4.3., 2.4.3.3., and 2.4.4. in the Method below). Note that although in previous research (Iordanou, 2010; Kuhn et al., 2008; Kuhn et al., 2013; Felton & Kuhn, 2001) participants had access to relevant evidence and were encouraged verbally to use evidence, no reflective activities focusing on evidence usage were employed. In addition to explicit reflective activities, following on the study by Kuhn et al., (2008), other techniques employed to heighten meta-level awareness of discourse were to conduct dialogues on the computer and to have participants to work in pairs. Arguing on the computer, through instant messaging software, has the benefit of providing an immediately available, permanent record of the discourse for participants to reflect on, in contrast to the conditions of real-time verbal discourse. In addition, having participants who held the same position collaborating against a series of opposing pairs who held the opposite position, we hypothesized, would make the collaborative pair more aware and reflective regarding the discourse and the use of evidence.

Furthermore, the present study employed engagement in peer discourse that prior work (Kuhn, Shaw, & Felton, 1997; Schwarz, Newman, & Biezuner, 2000; Chinn, 2006; Nussbaum & Sinatra, 2003) showed to be a facilitative method for the development of argument skills. According to Graff (2003), dialogic argument offers the “missing interlocutor” that is missing in individual argument. Students were particularly asked to persuade their interlocutors, based on Berland and Reiser’s (2010) finding of more engagement in evaluating and critiquing other’s position when students were arguing with the goal to persuade rather than to make sense. We have also incorporated engagement in a goal-based practice in dialogic argumentation in our methodology, since previous work has shown the effectiveness of this approach in facilitating participant’s ability to switch their focus from their own position to the other’s position (Kuhn et al., 1997), which is an ability that is required in developing evidence-based argumentation. Critiquing the evidence of the other’s position and offering evidence which detracts from the power of the other’s position, both of which lie at the heart of evidence-based argumentation, involve attention to the other’s position.

In the present study, pre-service teachers collaborated with a same-side peer in arguing against successive pairs of peers on the opposing side on the topic of Climate Change. Participants also engaged in reflective activities regarding the use of evidence in argumentation. Both dialogic and reflective activities were conducted in the context of a web-based learning environment. The environment included an IM software, an electronic knowledge base on the topic of climate change that participants could use for finding evidence to support their claims, and electronic templates that students used for their reflective activities.

1.4. Research Question – Hypothesis.

The present study addressed the following research question: Do enhancement of meta-level awareness of discourse, along with engagement in peer discourse, support the development of evidence-based argumentation? Drawing on the findings of previous work showing that participants who engaged in an intervention that supported meta-level awareness of argumentive strategies, along with engagement in peer discourse, were able to demonstrate gains in using powerful argumentive strategies (Kuhn et al., 2008) we hypothesized that heightening meta-level awareness of the use of evidence in discourse, through explicit reflective activities, along with engagement in peer discourse, would support participants' skill in using evidence in argumentation (Hypothesis). Participants' argumentation skills were assessed prior to and after the intervention, when they were engaged in an electronically conducted argument with a peer who held an opposing view on a socio-scientific topic. Another group of pre-service teachers, which was also assessed at the same times of year as the experimental group but studied about the role of evidence in science in the context of regular curriculum, served in a control condition. Participants' dialogue transcripts were analyzed in terms of the overall use of evidence, the amount of evidence per argument and per counter-argument, the function of evidence use and the accuracy of the evidence employed. We hypothesized that participants who engaged in peer discourse and in activities that aimed to heightened meta-level awareness of using evidence, would use overall more evidence, would use evidence both in their arguments and counter-arguments, would increase the use of accurate evidence and would employ evidence with the aim of satisfying the goals of skilled argumentation (Walton, 1989) of addressing and weakening the opponent's position.

2. Method

2.1. Participants

Participants were 66 Junior (third-year) students who attended an undergraduate program in Education in a public medium-sized state university in a city in South-East Europe. They were on average 20 years old. The 39 – 34 female and 5 male − participants in the experimental condition comprised the entire class of undergraduate students who attended a course in science education in a particular year. The 27 – 25 female and 2 male − participants in the control condition comprised the whole class of students who attended the same course the following year at the same university. Students were admitted to the program after they had passed competitive national entry exams. Participants came from a middle-class population. All students were enrolled in a required science education course and participated in the intervention as part of the course curriculum.

2.2. Learning Environment

A Learning Environment (LE) developed by Kyza and Constantinou was used for the purpose of the present study (see Figure 1). The LE was developed in the web-based learning platform of STOCHASMOS (Kyza & Constantinou, 2007). STOCHASMOS offers two main environments. The first environment was the Inquiry Environment, where a knowledge base for the topic of climate change was developed. The knowledge base included different types of information – short texts that described data, graphs, tables and images – which students could use to support their claims. Examples of data included in the knowledge base were a graph showing the atmospheric CO_2 levels from 1960 to 2010 from Wikipedia and a figure showing the surface temperatures over the last 1,100 years from the National Academy of Sciences. After each table, figure, or graph a short text followed which offered a simple description of it. The inclusion of this short description aimed to help students overcome any difficulty they might have had in interpreting data in different formats and therefore rule out the possibility that participants had difficulty in interpreting data in the case that they didn't use data adequately. Equivalent sets of data supporting each position (man-made vs. natural causes) were presented in the knowledge base. All experimental condition students had access to all the information presented in the knowledge base, supporting both man-made and natural causes.

The second environment is the WorkSpace environment, which hosts the reflective templates "Finding Evidence", "Own argument" and "Other argument" (see intervention section below), where students were asked to construct evidence-based arguments and reflect on the arguments they produced while they were engaging in dialogic argumentation. The XXX platform offered students the opportunity to transfer information from the Inquiry Environment to the WorkSpace environment, using the "Data Capture Tool". For example, for completing the "Other argument" reflective template students were asked to use their dialogue transcript to identify and report (by copying and pasting with the help of the Data Capture Tool) an argument that was offered by the opponent as well as the counter-argument they had offered to this argument. In a separate space in the template, the students had to report the evidence they used, if they did so, to support their counter-argument. The template then prompted students to consider revising their counter-argument. For revising their counter-argument, the template explicitly asked the students to report the evidence they had used to support their revised counter-argument. The platform also incorporates a chat tool, which was used for students' dialogic argumentation.

An initial version of the LE had been pilot tested using high school fifth graders (16-year-old students). After pilot testing, the learning environment had been revised based on students' and teachers' feedback, as well as our own observations. The revisions conducted concerned the content of the knowledge base. In particular, we shortened lengthy texts and added explanations for terms that students were not familiar with in an electronic glossary in the LE. Minor revisions, concerning the presentation of content knowledge, were also conducted after the implementation of the revised version of the learning environment with a group of high school 11^{th} graders.

Figure 1

2.3. Initial Assessment

Participants' argumentation skills were assessed through both initial and final assessments. Two socio-scientific topics, climate change (intervention topic) and alternative sources for generating electricity (transfer topic), were used for assessment. Control group participants underwent assessments on the transfer topic at the same times of year - with the same time interval between initial and final assessment - as the experimental group.

2.3.1. Individual argument. Initial positions and supporting arguments regarding the intervention topic − Climate Change (CC) − and the transfer topic − sources for generating electricity (SGE) – were assessed individually in writing. The scenario regarding climate change was based on the scenario developed by Sadler, Chambers and Zeidler (2004). The scenario presented two scientists who supported two alternative interpretations of climate change. One scientist supported that climate change is a natural phenomenon. He based his position on data from the history of the earth's climate showing the dynamic nature of climate and on climatic data collected by satellites, revealing no long term temperature increases caused by human activity. Another scientist supported that climate change is man-made. He based his position on recently recorded data showing an increase in atmospheric heat as a direct result of greenhouse gas accumulation. The latter scientists believed that human activities, mostly the burning of fossil fuels like oil, gas and coal, had significantly increased the concentration of atmospheric carbon dioxide since the middle of the 20th century. Participants indicated their position by choosing among the following options: ''Climate Change is Man-made", "Climate Change is Natural" and ''Undecided.'' The scenario for the transfer topic was about two alternative sources for producing electricity (see Appendix A). The scenario for the transfer topic included a short list of some relevant facts, three pieces of evidence for each position. The reason for providing this constrained knowledge base was to make sure that students had some data available if they wished to use them and also to rule out the possibility of lack of topic knowledge in case that they wouldn't use any data in their dialogues. No particular instructions were provided to students for using this information.

Response options for this topic were: "Coal should be used", "Natural Gas should be used" and "Undecided". For each scenario, participants were also asked to indicate the certainty of their position on a 6-point Likert scale, with endpoints labeled ‘‘totally certain’’ and ‘‘totally uncertain.’’ They were then asked for reasons supporting their position and finally for reasons that would support the opposing view.

2.3.2. Dialogic argument with opposing-view partner. Participants in the experimental condition engaged in an electronically conducted argumentation on the topics of CC (intervention topic) and SGE (transfer topic) with a partner holding an opposing view. Participants in the control condition engaged in an electronically conducted argumentation on the topic of SGE. For each topic, two groups were formed. For the climate change topic, one group consisted of participants who considered that the cause of climate change was man-made and the other group consisted of participants who considered that climate change was natural. For the electricity generation topic, one group consisted of participants who were in favor of coal as a source of electricity generation and another group consisted of participants who were in favor of natural gas. The assignment was based on the position statement expressed on the 6-point opinion scale and the reasons participants offered for supporting their position and the opposing view at the initial individual assessment (see 2.3.1 Individual Argument). The participants who indicated they were undecided gave reasons on both sides of the issue, man-made and natural for the CC topic and natural gas and coal for the SGE topic, and were assigned to one or the other position in a way that served to equate the number of participants on each side. In particular, we assigned all students who chose the undecided option at the individual assessment to the natural position for the CC topic and to the coal position for the SGE topic, which were the least popular sides. Prior to assigning undecided students to a particular side, we asked them if they could argue in favour of this particular side and all of them expressed their willingness to do so.In the experimental condition, for the CC topic, one group consisted of 20 students who chose the man-made position; another group consisted of 4 participants who chose the natural position and 15 who were undecided. For the SGE topic, one group consisted of 20 students who were in favour of the natural gas position and another group consisted of 3 students who were in favour of the coal option and 16 who were undecided. Similarly, in the control condition two groups were formed for the SGE topic. One group consisted of 14 participants who were in favour of the natural gas position and another group consisted of 7 students who were in favour of the coal position and 6 who were undecided.

By drawing one from each of the contrasting groups, pairs of opposing-side participants were formed for each topic. Pairs were different for the two topics. Each participant discussed with the same opposing partners at initial and final assessments. Because the total number of participants was an odd number in both the experimental condition group (N = 39) and the control condition group (N = 27), two participants from the experimental condition, one for each topic, and one participant from the control condition, engaged in electronic argumentation twice, at both initial and final assessment, so that all participants had the chance to argue with a partner holding an opposing position.

Dialogues were conducted via the XXX chat tool. The instructions given to participants were to try to convince their partners that their position was the better one. These dialogues lasted 20 minutes or less, if participants indicated they had finished. The software saved the dialogue for later analysis.

2.4. Intervention

Students in the experimental condition engaged in an intervention that took place once a week, for eight weeks, as part of a mandatory undergraduate course. Students were randomly assigned to three groups of 13, 13, and 14 students for practical reasons (due to a restricted number of laptops we had available to work with). Students in all groups participated in exactly the same curriculum as described below. Figure 2 shows the sequence of the activities that students in the experimental condition engaged in.

2.4.1. Session 1. Introduction to the mission: Extreme weather conditions. The first session had a twofold aim, which was to introduce participants to the project and their mission, and to get them familiar with the learning environment. In this unit, participants studied information about extreme weather conditions and filled out the first template, which consisted of two sections. In the first section participants had to describe an extreme weather phenomenon observed in our days and in the second one they had to provide evidence – information showing the existence of the phenomenon. Participants used a different template for each phenomenon they described. While participants were working on this template, researchers discussed with them, in person, the importance of using evidence. The metaphor of a tree was used for this purpose, comparing the claim of an argument with the branches of a tree and the evidence supporting the claim with the trunk of a tree.

In the first session, participants were also introduced to the learning environment of XXX and had the opportunity to practice using the capture tool, which enabled them to copy information from the learning environment and paste it in the workspace for filling in their templates.

2.4.2. Sessions 2 & 3. Conceptual Understanding: The Greenhouse effect. In the second and third session students studied the mechanism of the Greenhouse effect. They studied a description about how the mechanism works − included in the second unit of the learning environment of XXX − and performed the Herschel experiment.

2.4.3. Sessions 4, 5 & 6. Finding evidence & Argumentation: Is climate change man-made or natural? Same gender pairs were formed within each team (see section 2.3.2. on how teams where formed). The same-side pairs remained together until the showdown preparation.

2.4.3.1. Finding evidence. In the fourth session participants initially discussed the importance of using evidence to support their claims. They were asked to prepare evidence-based arguments that they could use during the electronic chats that would follow. To do so, participants studied information about climate change, included in the third unit of the learning environment, and filled in the template “Finding evidence”. In this template participants had to state a claim supporting their position - either for the man-made position or the natural position - and provide evidence to support their claim. Participants were encouraged to prepare several evidence-based arguments, using different templates for each argument. The preparation of evidence-based arguments lasted for fifty minutes. Coaches facilitated each group’s work helping to keep the students focused on their task when necessary; they also provided support with technical issues relating to the learning environment and answered students’ questions without offering any direct instruction or feedback relating to content.

2.4.3.2. Electronic dialogues. In the same session, each pair conducted an electronic dialogue with another pair who held the opposing position on the topic. Oral instructions provided to each pair were to collaborate with their partner to determine what they wished to say and, when they reached agreement, to enter their response and send it to the opposing pair. Three coaches circulated, and when asked, helped students with technology issues. Dialogues lasted an average of 20 minutes.

2.4.3.3. Reflective analysis. Right after each electronic dialogue, pairs analyzed an electronic transcript of their dialogue, which was available on their laptops, using the templates ‘‘Other Argument’’ and ‘‘Own Argument’’. With the help of the ‘‘Other Argument’’ template the pair’s task was to analyze the opposing side’s argument and reflect on the effectiveness of the counter-argument they had made, focusing particularly on the evidence they used to back up their claims. They were then asked to consider possible improvements to this counter-argument, in respect to using appropriate evidence. With the help of the ‘‘Own Argument” template, the pair’s task was to review and evaluate their rebuttals to the counter-arguments offered by the opposing side to their own arguments and consider possible improvements. When some pairs finished the reflective analysis of their own dialogue’s transcript, they exchanged reflection sheets with other pairs, by exchanging laptops, to provide and receive feedback. The coaches facilitated the process by prompting participants to think of whether their responses to the opponent were as well grounded and effective as they could be.

In the fifth session, participants engaged in the same activities as in session 4, with some differences in time allocation. They engaged in finding evidence for 20 minutes, in electronic dialogic argumentation for 25 minutes, in reflective activities for 20 minutes and in another electronic dialogue, again for 25 minutes. Participants engaged in electronic dialogue with a different pair of students each time.

Session 6 involved an electronic dialogue, reflective activities and preparation for the final showdown that would follow.

2.4.4. Session 7. ‘‘Showdown’’ preparation session. Participants who had been working together as a pair during electronic argumentation and reflective activities separated and were assigned to two different preparation teams. One team was assigned to be ‘‘own argument’’ specialists, who worked on preparing evidence-based rebuttals, and the other was assigned to be ‘‘other argument’’ specialists, who worked on preparing evidence-based counter-arguments. Each preparation team had a coach who facilitated the group process. Both groups were told that the purpose of this session was to prepare for the impending ‘‘showdown.’’ The ‘‘own argument’’ specialists were told that their task was to become familiar with the possible counter-arguments the opposition might assert and to prepare evidence-based rebuttals to use in the showdown. The team created a set of ‘‘own argument – counter-argument – evidence-based rebuttal’’ sequences that were recorded onto colour-coded cards, distinguishing each part of the argument sequence. A separate colour card for recording pieces of evidence supporting the claim presented in the rebuttal was attached to the rebuttal card. The reason of including this separate colour card was to highlight the need to include evidence to support their critique. The reflection sheets completed in previous sessions were made available for this activity and further possible improvements were considered. Members of the other team were the ‘‘other argument’’ specialists. Their task was to review effective evidence-based counter-arguments to use when faced with opponents’ arguments. The cards produced by this team reflected the argument sequence of ‘‘other argument – evidence-based counter.’’ Again, a separate colour card for recording pieces of evidence supporting the claim presented in the counter-argument was attached to the counter-argument card. The reflection sheets were made available for this activity and further possible improvements were considered.

2.4.5. Session 8. ‘‘Showdown’’. In order to offer students the opportunity for greater participation, two showdowns took place. Participants on each side of the issue were divided into four teams - Teams A, B, C and D. The previous ‘‘specialists’’ (own argument and other argument) were represented equally on each team. Each team’s participants were seated in different rooms. Two electronic dialogues took place – the first between teams A and B, and the second between teams C and D – through the LE’s chat tool. The dialogue in each room was projected onto a wall screen in the room. All members collaborated to come to an agreement on the text to be sent to the opposing side. One member of each team was designated as typist. The showdown thus consisted of two electronic dialogues between the two opposing sides of approximately 40 minutes duration each.

2.4.6. Judging and feedback. The electronic dialogues produced in the showdown were represented in an argument map prepared by the researchers. Different columns appeared for each team, with their contributions arranged in order of occurrence from top to bottom. All statements were represented and connected with lines to show their interrelation. Different colours were used to label statements as effective, ineffective, or neutral argumentative moves. A point system was also applied, making it possible to declare a winning team. The argument map and associated point scoring were presented to participants in a session following the showdown.

Figure 2

2.5. Control Group

Students in the control condition engaged in the regular curriculum of an undergraduate course in Science Education. They participated in the same number of eight sessions, over the same time period of eight weeks, as the experimental condition students, but a year later. In particular, students in the control condition examined the topics of the nature of science, and epistemic beliefs and their influence on science education for two sessions and how to design effective inquiry activities for elementary school students, taking into consideration different learning theories and children's' initial ideas of specific scientific concepts, for six sessions. Students attended lectures and engaged in whole class and small group discussions. The same research assistants who were available to support the experimental condition students, were also available to support the control condition students. The research assistants, except the authors, were blind of the condition of the participants. Students, in the context of studying epistemic beliefs, studied about the central role of evidence in knowledge construction and evaluation. However, unlike experimental condition students who engaged in a series of peer discourse on a specific topic and reflective activities, control condition students studied about the importance of evidence use and evaluation in science, by studying and discussing theoretical and empirical scholarly articles and books. Note that students in the control condition studied about the use of evidence in the context of reviewing epistemic beliefs, as part of their regular curriculum; no specific instructions were provided to the instructor of the control condition group to modify the regular curriculum of the Science Education course.

2.6. Post-intervention Assessment

In the final assessment, participants engaged in a single computer-mediated dialogue with the same partners as in the initial assessment. As described under ‘‘Initial Assessment’’ experimental condition participants engaged in an electronic dialogue on both the CC (intervention) and SGE (transfer) topics, while control group participants engaged in an electronic dialogue on the SGE topic.

2.7. Coding

The dialogues were segmented into idea units, with an idea unit defined as an assertion with any possible accompanying justification that might have been provided. Then idea units that contained evidence were coded based on evidence's function using an adapted version of the coding scheme developed by Moore and Kuhn (under review). A statement was considered as evidence if it offered an implicit answer to the question "how do you know", and if the answer came from an available source (rather than someone's mind as in an opinion or an assumed fact). Each evidence-utterance received two codes. The first code categorized each piece of evidence based on its function. The second code categorized each piece of evidence based on its accuracy.

Categorization of evidence based on its function was differentiated on whether function applied at the procedural level or at the meta-level.

2.6.1. Using evidence at the procedural level. Idea units containing evidence that were used in the process of argumentation – procedural level – as reasons, were further coded into four categories based on its function. Category E1 was used for evidence employed to weaken opponent's claims and category E2 was used for evidence employed to support one's own claim. Note, that E2 codes were assigned when the focus of the use of evidence was on one's own position, i.e. participants used evidence to support their own position without paying attention and addressing the opposing position. On the other hand, when the focus of the use of evidence was on opponent's position, E1 codes were employed. The use of evidence to weaken opponent's position (E1) is an essential skill to be mastered if one is to engage in effective argumentive discourse. Its use reflects awareness of the dual goals of argumentive discourse noted earlier (to weaken the opponent's argument, as well as to gain concessions from the opponent that will support one's own argument). Category E3 was used if one intentionally reasoned against own side with evidence and category E4 was used if one intentionally supported others' side with evidence.

2.6.2. Using evidence at the meta-level. Idea units containing evidence that were used as a way to manage argumentation at the meta level, were categorized as either E5 or E6. Idea units that requested evidence from opponent were coded as E5, while idea units that critiqued opponent for not giving evidence or for having weak or incorrect evidence were coded as E6. Examples of all categories used to code evidence use based on evidence function are presented in Table 1.

Table 1.

The second code that each idea unit containing evidence received concerned the accuracy of the evidence. Five categories were used to code accuracy. C1 was used for accurate general knowledge, C2 was used for undocumented evidence claims, C3 for unsupported argument claims, C4 for distorted use of evidence available in the learning environment or in the scenario, and C5 for correct use of evidence from the learning environment or the scenario. Examples of all categories used to code evidence use based on its accuracy are presented in Table 2. For C5 category, we made a further distinction between vague – C5v − and specific − C5s − reference to appropriate data from the learning environment. We used C5s for the cases when participants made a specific reference to the source of the data (e.g. "as you can see in graph 5….") or provided detailed numeric data when available. For example, when participants provided evidence regarding the value of coal, evidence such as the following "coal's value corresponds to 1/10th of the price of other raw materials" was coded as C5s, whereas evidence such as "coal is the most economical option" was coded as C5v. Note that codes that received an E5 or E6 (justification or criticism of the quality of evidence given) did not receive a C code.

Two coders – the first author and another researcher – engaged in the coding of the dialogues. 15% of the dialogues were used for training. Another twenty-five percent of the dialogues were randomly selected and used to calculate inter-rater reliability. All dialogues were coded while coders were blind to the time, condition and identity of the participants. The Cohen's kappa for inter-rater reliability for coding idea units containing evidence was .87. After establishing inter-rater reliability, the remaining electronic dialogues were segmented and coded by one of the raters, again blind to time and identity of the participants.

Table 2.

3. Results

We sought to examine whether engagement in peer discourse and activities, which aimed to heighten meta-level awareness of using evidence, would support participants' skill in using evidence in argumentation, as we have hypothesized. Hence, we examined the effects of overall use of evidence, the amount of evidence per argument and per counter-argument, the function of using evidence and the accuracy of evidence used at initial and final assessment, on the intervention and transfer topic. The analysis was based on the 104 electronic dialogues produced at initial and final assessments by experimental condition and control condition participants. For participants who engaged twice in an electronic argumentation on a particular topic and time in order to have all students in pairs - given that the overall number of participants in both conditions was an odd number - we included in the analysis only their performance on the first dialogue they participated in. Two participants in the experimental condition were absent during the assessment of the transfer topic, one during initial assessment and one during the final assessment, and were excluded from the analysis of the experimental condition on the transfer topic.

3.1. Evidence Usage at Initial and Final Assessment on the Intervention Topic

A repeated-measures analysis of variance (ANOVA) was conducted to examine whether there was any significant change in experimental condition's performance in using evidence from initial to final assessment. Because there was a difference in the number of idea units produced at initial and final assessment, percentages of usage were calculated for each participant, rather than frequencies. An arcsine transformation was performed to normalize these proportions.

3.1.1 Overall usage of evidence. Before turning to the main analysis, based on the function and accuracy of evidence, a quantitative indicator was examined, which while hardly conclusive on its own, provides an initial, albeit superficial, indication of change in evidence usage during argumentation. This is the change in the proportion of idea units which contained evidence from initial to final assessment. The average of idea units containing evidence increased from 56% (SD = 18.45) at initial assessment to 74% (SD = 11.68), at the final assessment, $F(1, 38) = 25.54$, $p < .001$, partial η^2 =.402.

3.1.2 Evidence per argument and per counter-argument. The amount of evidence per argument and per counter-argument was assessed to examine whether there was a change in the usage of evidence to support students' own position and to weaken opponents' arguments, from initial to final assessment. A repeated measures ANOVA showed that there was a significant increase from initial to final assessment in both the mean number of evidence per argument, $F(1, 38) = 13.825$, $p = .001$ partial $\eta^2 = .267$ from .46 (SD = .31) to 1.38 (SD = 1.16), and the mean number of evidence per counter-argument, from .51 (*SD* = .44) to 1.93 (*SD* = 1.08), $F(1, 38) = 77.406$, $p < .001$, partial $\eta^2 = .671$.

3.1.3 Evidence Category. We examined the purpose for which participants employed each piece of evidence. In other words, what was the function of evidence in the dialogue? We examined evidence function at the procedural level − using pieces of evidence in the process of argumentation as reasons – and at the meta-level − using evidence as a way to manage argumentation. Tables 3 and 4 show descriptive data, mean frequencies and mean percentages of usage of evidence, respectively, by evidence category and Time for experimental and control condition participants.

3.1.3.1 Using evidence at the procedural level. A repeated measures ANOVA – with Time as the within time variable – was conducted for each of the evidence categories E1 to E4.

Analysis concerning category E1 − evidence functioned to weaken opponents' claims – revealed that participants increased significantly the usage of this category from initial to final assessment. In particular, at initial assessment the mean percentage of evidence used to weaken opponents' claims was 18.96% (*SD* = 16.36), whereas at the final assessment the corresponding percentage was nearly twice as big: 35.46% (*SD* = 16.10), $F(1, 38) = 22.18$, $p < .001$partial η^2 =.369.

No significant change was observed in the usage of the other categories of evidence (E2, E3 and E4). As seen on Table 4, the mean percentage of usage of categories E3 and E4 was below 5% at initial assessment and this percentage remained at the same low level, around 5%, at the final assessment. It is noteworthy that most of the evidence used at initial assessment functioned to support a claim − category E2 − 24.70% (*SD* = 17.86). The percentage of idea units containing evidence that were categorized as E2 remained high at the final assessment, 20.48% (*SD* = 13.02), although the percentage of usage of E2 category was not the highest one at the final assessment.

3.1.3.2 Using evidence at the meta-level. Criticizing opponents for not giving evidence or for having weak or incorrect evidence (E6) was the other category in which a significant increase was observed from initial to final assessment. At initial assessment the mean percentage of usage of this category of evidence was 2.03% (*SD* = 4.75), whereas at the final assessment the corresponding percentage increased significantly to 5.52% (*SD* = 6.30), $F(1, 38) = 10.82$, $p = .002$, partial $\eta^2 = .222$. Although participants' engagement in meta-level talk about evidence doubled from initial to final assessment, the percentage of engagement in meta-level talk at the final assessment was still very low. We discuss this finding further in the discussion section.

Table 3.

Table 4.

3.1.4 Individual patterns of change. In addition to analyses of group trends, equally informative are analyses of changes at the individual level. The percentage of participants who produced at least one piece of evidence or at least three pieces of evidence was examined for each evidence category. The criterion of ‘‘at least three’’ ensures that use of evidence was not a random incident, but the result of significant mastery of the skill. Analysis at the individual level confirmed the pattern observed in group analysis. A significant increase was observed in the number of participants who used evidence to weaken opponents' claims (E1) and to criticize opponents for not giving evidence or for having weak or incorrect evidence (E6). At initial assessment, 77% − 30 out of 39 – of the participants used evidence category E1 at least once, whereas at the final assessment all of the participants did so (*p* =.021, McNemar test).

Concerning evidence category E6, 23% − 9 out of 39 – used this evidence category at least once at initial assessment, whereas the corresponding percentage at the final assessment was almost two times the percentage observed at initial assessment, 41% – 16 out of 39 – (p = .016, McNemar test).

Applying the most strict criterion of "at least three", 33% −13 out of 39 – used evidence category E1 at initial assessment, whereas 82% − 32 out of 39 − did so at the final assessment (p < .001, McNemar test). Regarding evidence category E6, none of the participants used this category at least three times at initial assessment and only one did so at the final assessment.

No significant change was observed in the usage of the other categories of evidence from initial to final assessment, applying either the "at least one" criterion or the "at least three" criterion. Whereas only few participants used evidence categories E3 and E5 at either initial or final assessment, a considerable number of participants used evidence for supporting own claims (E2) at both initial and final assessment. In particular, at initial assessment 36% − 14 out of 39 – and 77% − 30 out of 39 – of the participants used evidence category E2 at least three times or at least once, whereas at the final assessment 33% − 13 out of 39 – and 82% − 32 out of 39 − of the participants did so, correspondingly.

3.1.5 Accuracy of evidence. Besides examining the function of evidence, we also examined the accuracy of the evidence used. In this analysis we included only the idea units which contained evidence. As seen in Table 5, participants produced significantly less unsupported argument claims, evidence category C3, at the final assessment compared to the initial assessment. In particular, of the idea units containing evidence, the percentage of unsupported argument claims produced at the final assessment – 6.41% (*SD* = 10.36) − was one third of the corresponding percentage produced at initial assessment – 19.91% (*SD* = 21.80), *F*(1, 38) = 8.77, *p* =.005, partial η^2 = .188. Another change, not significant but in the same direction, is the reduction of the percentage of usage of undocumented evidence claims, that is evidence category C2. The percentage of usage of evidence category C2 was 11.58% (*SD* = 12.85) at the final assessment, which has reduced to half of the corresponding percentage produced at initial assessment – 20.60% (*SD* = 23.58).

Regarding correct use of evidence, which is evidence category C5, we observed some very interesting findings. Participants reduced vague reference to evidence, which is category C5v, at the final assessment – 7.45% (*SD* = 12.74) – compared to the initial assessment – 22.84% (*SD* = 25.97) − *F*(1, 38) = 8.32, *p* =.006, partial η^2 = .180. While the percentage of usage of vague reference to data decreased, the percentage of usage of specific reference to appropriate data from the learning environment increased significantly. Participants' percentage of specific reference to appropriate data at initial assessment, as seen in Table 5, was 33.17% (*SD* = 30.69), while at the final assessment it was 71.96% (*SD* = 19.95), *F*(1, 38) = 42.63, *p* <.001, partial η^2 =.529.

3.2 Evidence Usage at Initial and Final Assessment on the Transfer Topic

A two way repeated-measures analysis of variance was performed, 2 X 2 (Condition X Time) − with Time as the within subject variable − to investigate whether there were any differences between control and experimental group participants' performance from initial to final assessment on the topic of SGE regarding overall evidence usage, the amount of evidence per argument and counter-argument, the function of evidence and the accuracy of evidence employed.

3.2.1 Overall Evidence usage

An analysis of overall evidence usage revealed a 2-way significant interaction, $F(1, 62) = 8.75$, $p = .004$; partial $\eta^2 = .124$, for Condition X Time[1]. Overall the percentage of idea units containing evidence increased in the experimental condition from initial to final assessment − from 54.59% ($SD = 18.66$) to 66.22% ($SD = 17.22$) −, while no change was observed in the performance of students in the control condition – 44.84% ($SD = 14.37$) at initial assessment and 40.19% ($SD = 22.26$) at the final assessment.

3.2.2 Evidence per argument and per counter-argument

A 2 X 2 (Condition X Time) repeated-measures ANOVA on the amount of evidence per argument revealed a significant Time X Condition interaction, $F(1, 62) = 10.256$, $p = .002$, partial $\eta^2 = .144$. Experimental condition students increased the mean number of evidence per argument from .68 ($SD = .62$), at initial assessment, to 1.49 ($SD = 1.08$), at the final assessment. No significant change was observed in the control group's performance from initial, 1.11 ($SD = .78$) to final assessment .82 ($SD = .61$).

A two way ANOVA on the amount of evidence per counter-argument also revealed a significant Time X Condition interaction, $F(1, 62) = 44.660$, $p < .001$, partial $\eta^2 = .423$. Experimental condition students increased the amount of evidence per counter-argument from initial, .86 ($SD = .42$), to final assessment, 2.14 ($SD = .87$), while control condition students did not exhibit any significant change from initial, 1.41 ($SD = .92$) to final assessment, .92 ($SD = .80$).

3.2.3 Evidence Category

[1] A two-way ANOVA (Condition X Time) using the dyad as the unit of analysis was also significant for overall evidence usage, $F(1, 31) = 5.12$, $p = .031$, partial $\eta^2 = .142$.

3.2.3.1 Using evidence at the procedural level

A 2 X 2 (Condition X Time) repeated-measures ANOVA for evidence functioning to weaken opponent's claims (E1) revealed a two-way interaction, $F(1, 62) = 6.56$, $p = .013$, partial $\eta^2 = .096$[2]. Experimental group participants' mean percentage of usage of evidence category E1 increased from 22.45% (*SD* = 14.25), at initial assessment, to 33.22% (*SD* = 12.93), at the final assessment, showing participants' ability to transfer skills developed in the intervention topic to a new topic. Control group's usage of evidence category E1 did not exhibit any change from initial – 17.94% (*SD* = 12.37) − to final assessment – 17.68% (*SD* = 17.11).

A two way ANOVA on each of the evidence categories E2, E3, E4 and E5[3] showed that no significant change could be observed from initial to final assessment. Experimental condition participants, as seen in Table 4, exhibited small proportions of usage of E3 and E5 evidence categories at both initial − 5.45% (*SD* = 6.79) and 1.45% (*SD* = 4.04) − and final assessment – 5.78% (*SD* = 6.41) and 2.25% (*SD* = 5.46), correspondingly. Control condition participants also exhibited small proportions of usage of E3 and E4 evidence categories − 2.15% (*SD* = 4.20) and 2.74% (*SD* = 4.88) at initial assessment and 0% and 2.66% (*SD* = 6.83) at the final assessment, respectively.

Noteworthy, a high percentage of idea units functioned to support a claim (E2). In particular, experimental group participants' percentage of E2 was 24.50% (*SD* = 11.31) at initial assessment and 21.78% (*SD* = 10.44), at the final assessment. Control group's corresponding

---

[2] A two-way ANOVA (Condition X Time) using the dyad as the unit of analysis was also significant for E1 category of evidence, $F(1, 31) = 6.04$, $p = .020$, partial $\eta^2 = .163$.

[3] The scores of E5 category were not normally distributed even after the arcsine transformation, so the results were conformed running non parametric tests.

percentages were 21.44% (*SD* = 9.44) at initial assessment and 19.48% (*SD* = 10.45) at the final assessment. No significant change was observed from initial to final assessment regarding E2.

3.2.3.2 Using evidence at the meta-level. A two-way ANOVA (Time X Condition) on evidence category E6 − using evidence at the meta-level− revealed a significant change for the experimental condition, $F(1, 62) = 13.28$, $p = .001$, partial $\eta^2 = .174$[4]. By the end of the intervention, experimental group participants increased the percentage of usage of evidence for criticizing opponents for not giving evidence or for having weak or incorrect evidence, from 1.23% (*SD* = 3.01) to 3.85% (*SD* = 4.18). No significant change was observed in the control group's performance in using evidence at the meta-level. Control group's percentage of usage of evidence at the meta-level, was 1.12% (*SD* = 2.94) at initial assessment and 0% at the final assessment.

3.2.4 Individual patterns of change

3.2.4.1 Category E1. Most of the participants produced at least one piece of evidence for weakening opponents' claims (E1) at both initial and final assessment. In particular, 87% − 32 out of 37 − of experimental condition students and 81.5% − 22 out of 27 – of control condition students used at least one piece of E1 evidence at initial assessment, whereas the corresponding percentages at the final assessment were 97% − 36 out of 37 – and 70.4% − 19 out of 27. Applying the most advanced criterion of producing "at least three" pieces of evidence of a particular category, only 57% − 21 out of 37 – of experimental condition participants produced at least three pieces of E1 category at initial assessment, whereas a significantly greater proportion – 84%; 31out of 37 − did so at the final assessment ($p = .002$, McNemar test). No significant

[4] A two-way ANOVA (Condition X Time) using the dyad as the unit of analysis was also significant for using evidence at the meta-level, $F(1, 31) = 9.92$, $p = .004$, partial $\eta^2 = .242$

change was observed in control condition's performance from initial − 48%; 13 out of 27 – to final assessment – 30%; 8 out of 27.

3.2.4.2 Category E6. The number of experimental condition participants who produced at least one piece of evidence to criticize opponent (E6) was small at initial assessment − 6 out of 37 – but increased significantly at the final assessment − 19 out of 37, ($p < .001$, McNemar test). The corresponding number of control condition participants was small at initial assessment (11%; 3 out of 27) and remained at about the same low level at the final assessment (4%, 1 out of 27). No significant change was observed in either the experimental or control condition participants' performance when the criterion of "at least three" was applied.

3.2.4.3 Category E2. Regarding using evidence to support own position (E2), all experimental condition participants produced at least one piece of this category of evidence at both initial and final assessment. When the more strict criterion of "at least three" E2 pieces of evidence was applied, a significant increase was observed in the number of experimental condition participants, from 62% –21out of 37 – at initial assessment, to 84% at the final assessment – 31 out of 37 – ($p = .002$, McNemar test). The majority of control condition participants also produced at least one and at least three evidence of E2 category, at both initial and final assessment, but no significant change was observed from initial to final assessment (from 27 to 23 and from 13 to 8, respectively).

3.2.5 Accuracy of Evidence

Experimental group participants at the final assessment decreased significantly the vague reference to appropriate data from the learning environment, while they increased the specific reference to data, replicating the finding observed on the intervention topic. In particular, as seen on Table 5, the proportion of evidence that made explicit reference to data from the scenario increased in experimental group condition from 40.41% (*SD* = 22.20) at initial assessment to 47.06% (*SD* = 18.47) at the final assessment, $F(1, 36) = 4.26$, $p = .046$, partial $\eta^2 = .106$. The proportion of evidence that made vague reference to data decreased from 22.85% (*SD* = 18.25), at initial assessment, to 14.58% (*SD* = 13.53), at the final assessment, $F(1, 36) = 5.45$, $p = .025$, partial $\eta^2 = .131$. No significant change was observed in the control group participants' use of evidence which made vague or specific reference to data, 32.15% (*SD* = 19.13) and 21.58% (*SD* = 15.70) at initial assessment and, 26.21% (*SD* = 28.39) and 18.11% (*SD* = 22.59) at the final assessment, respectively.

No other significant change was observed in the use of the other evidence categories in either experimental or control group performance.

Table 5.

4. Discussion

Our findings show that enhancing meta-level awareness of the use of evidence in discourse, along with engagement in peer discourse, promotes more consistent, more accurate and broader use of evidence in argumentation. The results of the study confirmed our hypothesis, in showing that participants who engaged in an evidence-focused dialogic intervention used overall more evidence, used more evidence both in their arguments and counterarguments, increased the use of accurate evidence and employed evidence with the aim of satisfying the goals of skilled argumentation (Walton, 1989) of addressing and weakening the opponent's position. Participants exhibited advancements in their argumentation skills in the intervention topic – climate change – and transferred these advancements to a new topic – sources for generating electricity. In particular, they used two times more evidence with the purpose to weaken opponent's claims at the final assessment than they did at initial assessment, on the intervention topic, and this advancement transferred to the new topic. In addition, participants exhibited advancements regarding the accuracy of the evidence used. Participants produced less unsupported argument claims at the final assessment, than they had done at the initial assessment. No advancements were observed in participants' use of evidence in another group of pre-service teachers which did not engage in dialogic and reflective activities, but were engaged in the regular curriculum, and served as a control condition.

The increases observed on the part of almost all participants in the experimental condition signal an importance change in participants' focus of attention, from initial to final assessment, when engaged in dialogic argumentation. Although participants initially devoted more attention to their own position, as was reflected in the high percentage of usage of evidence supporting one's own claims, by the end of the intervention, they shifted their focus of attention to the other's position. Most of the times that participants used evidence at the final assessment, they did so with the purpose to weaken other's claims. This finding of transferring attention from one's own position to the other's position after participants engaged in an intervention based on dialogic activities over an extended period of time is consistent with findings of previous research (Iordanou, 2010, 2013; Kuhn et al., 2008; Kuhn & Udell, 2003; Felton, 2004). Paying attention and addressing other's position is the central and most essential skill to be mastered if one is to engage in effective argumentive discourse (Kuhn et al., 2008). This shift of attention reflects awareness of the dual goals of effective argumentive discourse, which are to weaken the opponent's argument, as well as to gain concessions from the opponent that will support one's own argument (Walton, 1989). Addressing other's position is according to Graff (2003) what gives status and value to an argument itself. Addressing other's position, besides its centrality in effective argumentation skills, implies an important epistemological achievement, that is an understanding that alternative interpretations can account for the same issue and, even more importantly, an understanding that different interpretations are amenable to evaluation. Evidence for achieving a more constructivist view of knowledge after participants engaged in discourse with peers who held an opposing view on the topic comes from the study of Iordanou (2010). In addition, we believe that the ability to address other's claims, especially when one addresses the opponent's critique to his/her own position, might support better awareness and reflection on one's own position, which is essential for belief and knowledge revision. However, our focus in the present work was on evidence-based argumentation. The gains observed in evidence-based argumentation do not necessarily imply any conclusions regarding possible change in the influence of participants' prior beliefs. Further research is required to investigate the possible effect of engagement in the intervention employed in the present study on participants' prior beliefs.

The control condition participants' performance showing no improvement, points to the importance of paying specific attention to the development of argumentation skills. Students in the control condition read scholarly articles and participated in lectures, in whole-class and small group discussions about the importance of using evidence in science in the context of studying about the nature of science and epistemic beliefs for two sessions; even so, they did not show any improvement in the use of evidence when engaged in argumentation. This finding shows that the limited opportunities offered in the regular pre-service teachers' curriculum about evidence in science were not sufficient to develop the participants' appreciation of the role of evidence in scientific knowledge. On the other hand, the gains observed in the experimental condition students suggest that engagement in a meaningful educational activity, which is based on sustained practice in dialogic argumentation and reflection, in the context of an interactive learning environment, appears a promising way for the development of evidence based argumentation skills. Working in the context of a learning environment which has been designed for the purpose of this intervention, including topic specific information and the appropriate presentation of it − in a way that corresponds to the educational level of the participants − was a facilitating factor, we believe, for the development of the skill of producing evidence-based arguments. Working in the context of a carefully designed learning environment reduces the cognitive load that participants would otherwise have to deal with when freely surfing on the web, searching for information or when dealing with long complex texts, as we have experienced while pilot-testing the learning environment. The learning environment might have facilitated students to focus their attention and devote their cognitive resources on developing their skill of incorporating evidence in their claims.

The present work points to the importance of meta-level awareness and understanding of argumentation rather than simply the implementation of argument strategies and the use of evidence at the procedural level. In particular, our findings show that heightening meta-level awareness of the use of evidence in argumentation, through direct reflective activities regarding the use of evidence and by conducting dialogues on the computer while they were collaborating with a same-side peer, along with engagement in discourse, support the development of evidence-based argumentation skill (Hypothesis). Besides the gains observed at the procedural level, gains were also observed at the meta-level, supporting Kuhn's (2001) model, which proposed that development of skills at the procedural level is supported by development of meta-level knowledge. Participants in the experimental condition increased the percentage of claims they used to criticize the opponent for not giving evidence or for having weak or incorrect evidence at the final assessment compared to the initial assessment, revealing an improvement in meta-strategic knowledge. This criticism made by participants reveals that they have developed, at least implicitly, an understanding that using evidence to back up claims is an important goal of argumentation. They demonstrated an understanding that strong arguments are the ones that are supported by relative evidence and when their interlocutors failed to incorporate evidence in their claims, they identified this as a weakness. Our findings are in line with other studies showing that enhancing meta-level knowledge promotes argumentation skills (Kuhn et al., 2008) and reasoning skills (Zohar & Peled, 2008, Zohar & Ben David, 2008). In particular, Kuhn et al., (2008) found gains in meta-strategic knowledge, in understanding the relevance of the other person's position, along with gains at the performance level, of producing more counter-arguments and rebuttals. In a broader sense, understanding and appreciating the value of using evidence in an argument is an epistemological achievement (Sandoval & Millwood, 2008). Judging an argument on the basis of whether it includes data that support the theory in which it is contained reveals an appreciation of the process of coordinating theories with evidence as a means of knowing. Also, criticizing others' claims − and their accompanied evidence − reveals an understanding that data do not speak out by themselves. On the contrary, data are amenable to different interpretations, a fundamental understanding for epistemological development (Iordanou, under review). Furthermore, our findings suggest that dense engagement in argumentation in a specific context along with evidence-specific reflective activities, where participants construct an understanding about the role of evidence in argumentation, might be a more facilitative condition for promoting meta-level awareness of the role of evidence in argumentation, which, in turn, appears to support the use of evidence in argumentation at the procedural level, than engagement in an abstract theoretical examination of the role of evidence in science. This finding is consistent with Sandoval (2005) who view practice in argumentation as a fruitful means for developing epistemic beliefs. Further research is warranted to elucidate more clearly the suggestive advantages observed in this study of an evidence-focused argumentive intervention in a specific context over limited engagement in studying evidence in an abstract, general context on appreciation of the role of evidence and employment of evidence in argumentation.

4.1. Limitations

The present study employed a multifaceted and complex intervention which included various components, such as discussions with peers who held an opposing view on the topic, collaboration between peers who held the same position and reflective activities. The multifaceted nature of the intervention does not enable us to draw conclusions regarding which specific element or elements of the intervention had contributed to the gains observed in the participants' use of evidence. Including different control groups that will control for possible explanations would be required in future research to delineate this.

Also, the transfer topic had some similarities with the intervention topic. Given the similarities between the two topics, one could argue that increases in topic knowledge might have accounted for the gains observed in experimental condition. However, evidence against such an effect comes from a study by Iordanou and Constantinou (under review) in which a control group was included that studied the same topic using the same learning environment, for the same amount of time, as the experimental condition students but did not exhibit any gains in evidence-based argumentation skills. In addition, the data rich scenarios that were provided diminished the possibility that the control group's failure to show improvement was due to lack of topic knowledge. Finally, the advancements at meta-level knowledge, which transferred from the intervention to the transfer topic, showed that the observed advancements in evidence use cannot be attributed to mere gains in topic knowledge.

4.2. Educational Implications

The present study has important educational implications. Science education researchers and educational psychologists have argued for the importance of developing students' argumentation skills in the context of science education, since evidence-based arguments lie in the heart of science (Kuhn, 2010; Driver et al., 2000). Yet, at the same time previous research has repeatedly revealed that students' argument skills, particularly their ability in producing evidence-based arguments, are not optimally developed through conventional science teaching. This finding applies for students of different educational levels – ranging from elementary school students (Iordanou, 2010; Sandoval, 2005) to college students (Kelly & Takao, 2002). The present study shows that engagement in discourse and reflective activities about the use of evidence supports the development of evidence-based argumentation skills.

4.3. Implications for Science Teacher Education

The present findings also have important implications for science teacher education in particular. Previous research has shown that teachers – both in-service and pre-service – pose weak argument skills (Zohar, 2004; Zembal-Saul, 2002), showing that regular teacher education curriculum has little effect on promoting these skills. These findings, in conjunction with the fact that very little work has been done about teacher education in the field of argumentation (Zohar, 2008), point to the urgent need for studies to illuminate our understanding in that respect. The present study addressed the issue of developing argumentation skills in a sample of pre-service teachers and showed clearly that when teachers' argumentation skills receive specific attention, in the context of a special intervention designed to promote these skills through enhancement of meta-level awareness, they are amenable to development. Our findings are in line with other studies **(**Kollar et al.'s, 2014; Kopp & Mandl, 2011) that showed that providing scaffolding of argumentation skills, in the context of an argumentive focused intervention, is an effective condition to support the development of argumentation skills and specifically the skill of engaging in evidence-based argumentation.

Engagement in such intervention programs, we claim, has multiple benefits. Firstly, it offers teachers the opportunity to develop their own argument skills, a fundamental prerequisite for teachers in order to promote their students' argument skills. It is unrealistic to expect teachers to adopt practices for supporting their students' argument skills if they have not themselves developed strong argument skills. Secondly, teachers become familiar with new learning practices that promote argumentation. This would make it possible that they would apply these practices, since previous research reveals the close connection between teachers' familiarity with a specific teaching approach and the adoption of that particular approach in their practice (Zohar, 2004; Avraamidou & Zembal-Saul, 2005). Finally, engagement in methods like the one employed in the current study, which are based on argumentation and reflective activities, is a promising pathway for the development of teachers' understanding of the epistemological foundations of science, that is, recognizing scientific knowledge as constructed by humans rather than simply discovered in the world. This constitutes the major challenge for developing a mature epistemological understanding in the science domain (Kuhn et al., 2008).

Our findings, overall, suggest that the policy recommendation supported by previous research examining middle and high school students' argument skills − engagement and practice in argumentation within the context of authentic science topics – should also apply for the education of science teachers, if we want teachers to develop strong argument skills and promote an understanding that considers critical examination of evidence-based arguments as a practice at the heart of science.

References

Amsel, E., & Brock, S. (1996). The Development of evidence evaluation skills. *Cognitive Development*, *11*, 523-550. doi:10.1016/S0885-2014(96)90016-7

Anderson, R. C., Chinn, C., Chang, J., Waggoner, M., & Yi, H. (1997). On the logical integrity of children's arguments. *Cognition and Instruction*, *15*(2), 135-167. doi:10.1207/s1532690xci1502_1

Avraamidou, L., & Zembal-Saul, C. (2005). Giving priority to evidence in science teaching: A first-year elementary teacher’s specialized practices and knowledge. *Journal of Research in Science Teaching*, *42*, 965-968. doi:10.1002/tea.20081

Chinn, C. A. (2006). Learning to argue. In A. O’Donnell, C. Hmelo-Silver, & G. Erkens (Eds.), *Collaborative learning, reasoning, and technology*. (pp. 355-383). Mahwah, NJ: Erlbaum.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science education. *Review of Educational Research*, *63*, 1-49. doi:10.3102/00346543063001001

Chinn, C. A., Buckland, L. A., & Samarapungavan, A. (2011). Expanding the dimensions of epistemic cognition: Arguments from philosophy and psychology. *Educational Psychologist*, *46(3)*, 141–167, doi:10.1080/00461520.2011.587722

Crowell, A., & Kuhn, D. (2014) Developing dialogic argumentation skills: A 3-year intervention study. *Journal of Cognition and Development*, *15(2)*, 363-381, doi:10.1080/15248372.2012.725187

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84(3),* 287-312. doi:10.1002/(SICI)1098-237X(200005)84:3<287::AID-SCE1>3.0.CO;2-A

Erduran, S., Simon, S., & Osborne, J. (2004). TaPping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education*, *88(6).* 915-933. doi:10.1002/sce.20012

Felton, M. (2004). The development of discourse strategies in adolescent argumentation. *Cognitive Development, 19,* 35-52. doi:10.1016/j.cogdev.2003.09.001

Felton, M., & Kuhn, D. (2001). The development of argumentative discourse skill. *Discourse Processes, 32*, 135-153.

Franklin, A. (1994). How to avoid the experimenters' regress. *Studies in History and Philosophy of Science, 25(3),* 463-491. doi:10.1007/978-94-011-5334-8_2

Graff, G. (2003). *Clueless in academe: How schooling obscures the life of the mind*. New Haven, CT: Yale University Press.

Iordanou, K. (2010). Developing argument skills across scientific and social domains. Journal of *Cognition and Development. 11*(3), 293-327.

Iordanou, K. (2013). Developing face-to-face argumentation skills: Does arguing on the computer help? *Journal of Cognition and Development. 14*(2), 292-320.

Iordanou, K. & Constantinou. C. P. (2015). Supporting use of evidence in argumentation through practice in argumentation and reflection in the context of SOCRATES learning environment. *Science Education. 99,* 282–311.

Jiménez-Aleixandre, M. P., Rodriguez, A. B., & Duschl, R. A. (2000). "Doing the lesson" or" doing science": Argument in high school genetics. *Science Education, 84(6),* 757-792. doi:10.1002/1098-237X(200011)84:6<757::AID-SCE5>3.0.CO;2-F

Kelly, G. J., & Takao, A. (2002). Epistemic levels in argument: An analysis of university oceanography students' use of evidence in writing. *Science Education, 86(3),* 314-342. doi:10.1002/sce.10024

Kelly, G. J., Druker, S., & Chen, C. (1998). Students' reasoning about electricity: Combining performance assessments with argumentation analysis. *International Journal of Science Education, 20(7),* 849-871. doi:10.1080/0950069980200707

Kollar, I., Ufer, S., Reichersdorfer, E., Vogel, F., Fischer, F., & Reiss, K. (2014). Effects of collaboration scripts and heuristic worked examples on the acquisition of mathematical argumentation skills of teacher students with different levels of prior achievement. *Learning and Instruction, 32,* 22-36. doi:10.1016/j.learninstruc.2014.01.003

Kolstø, S. D., Bungum, B., Arnesen, E., Isnes, A., Kristensen, T., Mathiassen, K., . . . Ulvik, M. (2006). Science students' critical examination of scientific information related to socioscientific issues. *Science Education, 90(4),* 632-655. doi:10.1002/sce.20133

Kopp, B., & Mandl, H. (2011). Fostering argument justification using collaboration scripts and content schemes. *Learning and Instruction, 21(5),* 636-649. doi:10.1016/j.learninstruc.2011.02.001

Kuhn, D. (2001). How do people know? *Psychological Science, 12,* 1-8. doi:10.1111/1467-9280.00302

Kuhn, D. (2010). Teaching and learning science as argument. *Science Education, 94(5),* 810-824. doi:10.1002/sce.20395

Kuhn, D., Goh, W., Iordanou, K., & Shaenfield, D. (2008). Arguing on the Computer: A Microgenetic Study of Developing Argument Skills in a Computer-Supported Environment. *Child Development, 79*(1), 233-234.

Kuhn, D., Iordanou, K., Pease, M., & Wirkala, C. (2008). Beyond control of variables: What needs to develop to achieve skilled scientific thinking? *Cognitive Development, 23,* 435–451.

Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentative reasoning. *Cognition and Instruction*, *15,* 287-315. doi:10.1207/s1532690xci1503_1

Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development, 74,* 1245-1260. doi:10.1111/1467-8624.00605

Kuhn, D., Zillmer, N., Crowell, A., & Zavala, J. (2013). Developing norms of argumentation: Metacognitive, epistemological, and social dimensions of developing argumentive competence. *Cognition and Instruction,* 31(4), 456-496. doi:10.1080/07370008.2013.830618

Kyza, E. A., & Constantinou, C. P. (2007). *STOCHASMOS*: a web-based platform for reflective, inquire-based teaching and learning. [Computer software]. Cyprus: Learning in Science Group.

Lehrer, R., Schauble, L., & Petrosino, A. J. (2001). Reconsidering the role of experiment in science education. In K. Crowley, C. Schunn, & T. Okada (Eds.), *Designing for science: Implications from every day, classroom, and professional settings* (pp. 251-277). Mahwah, NJ: Erlbaum.

Maloney, J. F. (2007). Children's roles and use of evidence in science: An analysis of decision-making in small groups. *British Educational Research Journal*, *33(3)*, 371-401. doi:10.1080/01411920701243636

Maloney, J., & Simon, S. (2006). Mapping children's discussions of evidence in science to assess collaboration and argumentation. *International Journal of Science Education, 28(15),* 1817-1841. doi:10.1080/09500690600855419

Mason, L. (1998). Sharing cognition to construct scientific knowledge in school context: The role of oral and written discourse. *Instructional Science, 26(5),* 359-389. doi:10.1023/A:1003103213786

Mason, L., & Scirica, F. (2006). Prediction of students' argumentation skills about controversial topics by epistemological understanding. *Learning and Instruction, 16(5)*, 492-509. doi:10.1016/j.learninstruc.2006.09.007

Mercer, N., & Littleton, K. (2007). *Dialogue and development of children's thinking: A sociocultural approach.* London, United Kingdom; New York, NY: Routledge.

Moore & Kuhn (under review).

Muis, K. R. (2007). The role of epistemic beliefs in self-regulated learning. *Educational Psychologist*, *42(3)*, 173–190, doi:10.1080/00461520701416306

Nussbaum, E. M., & Sinatra, G. (2003). Argument and conceptual engagement. *Contemporary Educational Psychology, 28*, 384 -395. doi:10.1016/S0361-476X(02)00038-3

Osborne, G., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching, 41(10),* 994-1020. doi:10.1002/tea.20035

Petty, R. E., & Wegener, D. T. (1999). The Elaboration Likelihood Model: Current status and controversies. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 41-72). New York: Guilford.

Reznitskaya, A., Anderson, R. C., & Kuo, L. J. (2007). Teaching and learning argumentation. *The Elementary School Journal*, *107*(5), 449-472. doi:10.1086/518623

Sadler, T. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching*, *41*, 513-536. doi:10.1002/tea.20009

Sadler, T. D., Chambers, W. F., & Zeidler, D. L. (2004). Student conceptualizations of the nature of science in response to a socioscientific issue. *International Journal of Science Education*, *26(4),* 387- 409. doi:10.1080/0950069032000119456

Sandoval, W. A. (2005). Understanding students' practical epistemologies and their influence on learning through inquiry. *Science Education, 89,* 634-656. doi:10.1002/sce.20065

Sandoval, W. A., & Millwood, K. A. (2008). What can argumentation tell us about epistemology? In S. Erduran & M. P. Jiménez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 71-88). Dordrecht, The Netherlands: Springer. doi:10.1007/978-1-4020-6670-2_4

Schauble, L. (1990). Belief revision in children. *Journal of Experimental Child Psychology, 49,* 31-57.

Schwarz, B., Newman, Y., & Biezuner, S. (2000). Two wrongs may make a right . . . If they argue together! *Cognition and Instruction*, *18(4),* 461-494. doi:10.1207/S1532690XCI1804_2

Stanovich, K. E., West, R. F., & Toplak, M. E. (2013). Myside bias, rational thinking, and intelligence. *Current Directions in Psychological Science, 22(4),* 259-264. doi:10.1177/0963721413480174

Stein, N., & Miller, C. (1993). The development of memory and reasoning skill in argumentative contexts: Evaluating, explaining, and generating evidence. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 285–335). Hillsdale, NJ: Erlbaum.

Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511840005

Walton, D. N. (1989). Dialogue theory for critical thinking. *Argumentation, 3(2)*, 169–184. doi:10.1007/BF00128147

Zembal-Saul, C., Munford, D., Crawford, B., Friedrichsen, P., & Land, S. (2002). Scaffolding preservice science teachers' evidence-based arguments during an investigation of natural selection. *Research in Science Education*, *32*, 437-463. doi:10.1023/A:1022411822951

Zohar, A. (2004). *Higher order thinking in science classrooms: Students' learning and teachers' professional development.* Dordrecht, The Neatherlands: Kluwer Academic. doi:10.1007/978-1-4020-1854-1

Zohar, A. (2008). Science teacher education and professional development in argumentation. In S. Erduran & M. P. Jiménez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 245-268). Dordrecht, The Netherlands: Springer. doi:10.1007/978-1-4020-6670-2_12

Zohar, A., & Ben David, A. (2008). Explicit teaching of meta-strategic knowledge in authentic classroom situations. *Metacognition and Learning, 3(1),* 59-82. doi:10.1007/s11409-007-9019-4

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching, 39(1),* 35-62. doi:10.1002/tea.10008

Zohar, A., & Peled, B. (2008). The effects of explicit teaching of metastrategic knowledge on low-and high-achieving students. *Learning and Instruction, 18(4),* 337-353.doi:10.1016/j.learninstruc.2007.07.001

Appendix A

The scenario used for the transfer topic (SGE)

Households of a new Town have to decide on the kind of source they will be using for producing electricity in their new Town. A problem has come up.

A scientist, who was called to provide his viewpoint on the issue, claimed that Natural Gas should be used for producing electricity. “Natural gas is a gaseous mixture of hydrocarbons. Natural gas is colourless and odorless. The combustion of natural gas compared with that of other fuels such as coal or oil, has less harmful effects on the environment. For example, natural gas produces smaller quantities of carbon dioxide for every unit of energy produced than other sources, such as coal or oil. It is extracted from underground wells, where it is stored under high pressure. The gas has been formed in a way similar to the way of the formation of oil. The reserves of natural gas, although are limited at the moment compared to the reserves of coal, recently has increased due to the discovery of new sources of natural gas and in the future are likely to increase even more. The countries which have the biggest gas reserves are Russia, the United States of America, the United Arab Emirates, Canada and the United Kingdom. "

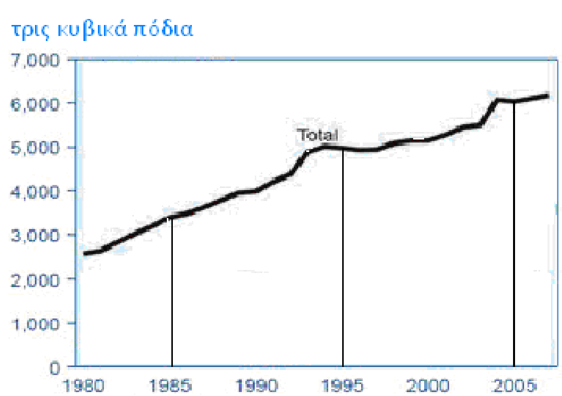
However, another scientist who was also called to provide his view about the issue claimed that Coal should be used for electricity production. “The Coal is a black rock used mainly as fuel. It is extracted from underground mines or open cavities in the ground. The countries with the biggest coal reserves are the United States, Russia, Australia, China, India and South Africa. Coal is the most economical option for electricity generation ─ its value corresponds to 1 / 10 of the price of other raw materials — and it's widely used in the industry because it contributes to the production of low cost industrial products. In addition, there are many coal reserves on Earth, which could last for many years. Although coal has been accused of releasing much more carbon dioxide than other materials when used for electricity production, scientists are studying recently ways for reducing the rates of carbon dioxide released during the process of electricity generation through coal."

Some possibly relevant facts

Natural Gas

- The cost of producing energy with natural gas is about 2.1 cents per kilowatt hour (0,021 € / KWh). The kilowatt-hour is a unit of energy.
- The production of 1 kilowatt of electricity releases 452 gr of carbon dioxide emissions into the atmosphere.
- The reserves of natural gas, based on the current volumes and consumption rates, are expected to last for about 50 years. However, new technologies recently, have discovered new gas volumes as seen in the graph 1 below. Graph 1 shows the reserves of natural gas from 1980 until a few years ago. Yet, no one can accurately predict if and how many stocks will be found in the coming years.

Graph 1. Reserves of natural gas.



Coal

- The cost of producing energy with coal is about 0.4 cents per kilowatt hour (0.004 € / KWh). The kilowatt-hour is a unit of energy.

- The production of 1 kilowatt of electricity releases 950 gr of carbon dioxide emissions into the atmosphere.

- The reserves of coal, based on the current volumes and consumption rates, are expected to last for about 115 years.

Table 1

Categories of Evidence based on their Function (E codes)

Category	Evidence Function	Examples
E1	To weaken other's claims	*Participant A*: Increase in carbon dioxide means increase in temperature *Participant B*: And what would you say about graph A5 which shows clearly that carbon dioxide concentration might increase, but the temperature stays stable?
E2	To support one's own claim	I think coal should be used in power stations because the cost of producing electricity using coal is much lower (0.4 cents per kilowatt hour) than natural gas (2.1 cents per kilowatt hour)
E3	To intentionally reason against own side with evidence	*Participant (who was in favour of natural gas):* Despite the fact that natural gas is expensive, …
E4	To intentionally support others' side with evidence	*Participant (who was in favour of coal):* The combustion of natural gas has less harmful effects on the environment.
E5	To request evidence from opponent	Do you have any evidence to support this?
E6	To criticize opponent for not giving evidence or for having weak or incorrect evidence	Probabilities…..! There is nothing certain; therefore you can't support your position using probabilities.

Table 2

Categories of Evidence based on their Accuracy (C codes)

	Evidence Accuracy	*Examples*
C1	Accurate general knowledge	fuel prices increase because of the economic crisis
C2	Undocumented evidence claims	the consequences of consumption of CO_2 will appear in 100 years from now
C3	Unsupported argument claims	it's reasonable and it's expected to have a rice in prices
C4	Distorted use of evidence	The cost of using coal is one third of the cost of using gas per kilowatt hour
C5	Correct use of evidence	The coal reserves will last for 115 years

Table 3

Mean Frequencies of usage of Evidence by Evidence Category, Condition, Topic and Time

	Experimental Condition (*N* = 39)								Control Condition (*N* = 27)			
	Intervention Topic				Transfer Topic				Transfer Topic			
	Initial Assessment		Final Assessment		Initial Assessment		Final Assessment		Initial Assessment		Final Assessment	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
E1	2.28	2.10	4.90	2.93	2.86	2.02	6.00	3.80	2.13	1.25	1.73	1.89
E2	2.92	2.94	2.46	2.49	3.39	1.87	3.81	2.12	2.50	1.11	1.90	1.60
E3	.41	.79	.62	.88	.72	.78	.94	.96	.20	.41	.00	.00
E4	.00	.00	.00	.00	.00	.00	.00	.00	.30	.54	.27	.69
E5	.54	.88	.90	1.07	.22	.59	.28	.62	.00	.00	.00	.00
E6	.33	.66	.51	.72	.19	.47	.72	.91	.15	.36	.00	.00

Table 4

Mean Percentage of usage of Evidence by Evidence Category, Condition, Topic and Time

	Experimental Condition (*N* = 39)								Control Condition (*N* = 27)			
	Intervention Topic				Transfer Topic				Transfer Topic			
	Initial Assessment		Final Assessment		Initial Assessment		Final Assessment		Initial Assessment		Final Assessment	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
E1	18.96%	16.36	35.46%***	16.10	22.45%	14.25	33.22%*	12.93	17.94%	12.37	17.68%	17.11
E2	24.70%	17.86	20.48%	13.02	24.50%	11.31	21.78%	10.44	21.44%	9.44	19.48%	10.45
E3	3.60%	5.29	4.10%	6.69	5.45%	6.79	5.78%	6.41	2.15%	4.20	0.00%	0.00
E4	0.00%	0.00	0.00%	0.00	0.00%	0.00	0.00%	0.00	2.74%	4.88	2.66%	6.83
E5	3.40%	9.14	5.50%	6.00	1.45%	4.04	2.25%	5.46	0.00%	0.00	0.00%	0.00
E6	2.03%	4.75	5.52%*	6.30	1.23%	3.01	3.85%**	4.18	1.12%	2.94	0.00%	0.00

*$p < .05$.

** $p < .01$.

***$p < .001$.

Table 5

Mean Percentage of Evidence Accuracy Categories in Idea Units Containing Evidence by Topic, Time and Condition

	Experimental Condition				Control Condition	
	Intervention Topic		Transfer Topic		Transfer Topic	
	Initial Assessment	Final Assessment	Initial Assessment	Final Assessment	Initial Assessment	Final Assessment
C1	0.00% (0.00)	0.00% (0.00)	0.00% (0.00)	0.00% (0.00)	3.68% (11.04)	1.31% (4.15)
C2	20.60% (23.58)	11.58% (12.85)	16.14% (18.48)	21.07% (13.53)	11.97% (13.69)	16.73% (23.96)
C3	19.91% (21.80)	6.41% (10.36)**	18.71% (15.92)	14.09% (12.29)	26.77% (20.95)	33.09% (32.22)
C4	3.47% (8.53)	2.61% (7.35)	1.71% (7.26)	0.50% (2.11)	3.80% (8.37)	3.06% (11.16)
C5s	33.17% (30.69)	71.96% (19.95)***	40.41% (22.20)	47.06% (18.47)*	21.58 (15.70)	18.11% (22.59)
C5v	22.84% (25.97)	7.45% (12.74)**	22.85% (18.25)	14.58% (13.53) *	32.15 (19.13)	26.21% (28.39)

*$p < .05$.

** $p < .01$.

***$p < .001$.