Scaffolding students in evaluating the credibility of evidence using a reflective web-based inquiry environment on Biotechnology

Iolie Nicolaidou, Eleni A. Kyza, Cyprus University of Technology,
P.O. Box 50329, 3603, Limassol, CYPRUS
Frederiki Terzian, Andreas Hadjichambis, Dimitris Kafouris, Cyprus Ministry of Education and Culture
Email: iolie.nicolaidou@cut.ac.cy, eleni.kyza@cut.ac.cy, freda.terzian@yahoo.com,
a.chadjihambi@cytanet.com.cy, dimitris_kafouris@yahoo.co.uk

Abstract: This case study investigated scaffolding to support twelve 11th grade students' collaborative construction of evidence-based explanations and their evaluation of the credibility of evidence through the utilization of a web-based reflective inquiry environment on Biotechnology. Over eleven 90-minute lessons students investigated and evaluated scientific data relating to the cultivation of genetically modified plants. The analysis of pre- and post-tests on students' conceptual understanding of Biotechnology topics and their skills in evaluating the credibility of evidence revealed learning gains and suggested that the intervention was successful. Students' written explanations in task-related artifacts and the analysis of two groups' videotaped discussions showed that the students became sensitive to credibility criteria, questioned the sources of data and correctly identified sources of low and high credibility. Students' difficulty in applying methodological criteria suggests that this criterion should be addressed in future studies.

Introduction

A central goal of science education is to support students' evidence-based reasoning by engaging them in seeking evidence and using it to critique scientific claims. According to the U.S. National Education Standards (NRC, 1996) students should be able to apply scientific reasoning to participate in informed decision-making at the local and national level about issues that impact on their everyday life, such as the global climate change and genetic engineering.   Participating in such decision-making requires an understanding of the nature of science, as well as scientific knowledge and skills relating to the interpretation and weighing of evidence, issues which are underrepresented in the current practice of teaching and learning science (Chinn & Malhotra, 2002). Reaching decisions on many socio-scientific debates that scientists and the public need to take action on requires the existence of processes such as making sense of complex and diverse data sets that are difficult to analyze and comprehend, and weighing the relevance and examining the credibility of scientific evidence. "Practices such as weighing evidence…and evaluating the potential viability of scientific claims are seen as (some of the) essential components in constructing scientific arguments (Driver, Newton & Osborne, 2000, p.288). Driver et al. (2000) noted that insufficient time is typically given to evaluative tasks beyond the interpretation of data; for instance, questions such as "what trust can we place in data?" or "are there different possible interpretations of this data?" are not frequently addressed. Indeed, often existing inquiry curricula have explicit or implicit expectations that students treat data as non-biased and do not raise concerns over the credibility of the evidence.

As Yang (2004) pointed out, given the complex issues associated with the use of evidence in ill-defined problems there is a need for more intensive investigation of students' potential to use multiple sources of evidence and to critically judge claims based on evidence. Students' engagement in solving complex socio-scientific problems provides a good outlet for having them evaluate the credibility of multiple pieces of evidence. Socio-scientific issues are real-life problems that bombard the citizens of modern society and frequently represent dilemmas, stemming from ethical aspects that need to be taken under consideration (Sadler, 2004) and from biased interpretations or presentation of the data.   As Gieryn (1999), a sociologist of science, argues "what happens in nature…depends upon the chef you ask: for some, nature is a seasoning thrown in to the flavor the social meat and cultural potatoes; for others, nature is what is finally brought to the table, what gets ladled into bowls, either thick stew with chunks of social left in or thin broth after the "meat" is methodically strained out and discarded; still others never bother to pick up any nature at the market –it is social down to the bottom of the pot." (Gieryn, 1999, p. ix.)

The work presented in this paper examines 11th grade students' assessment of the credibility of evidence and describes a research-validated pedagogical approach in supporting students' attending to such issues. This work contributes to the understanding of students' capacity to evaluate the credibility of evidence, given the need to prepare future citizens in critically evaluating the credibility of evidence and the limited discussions in the literature about students' evaluation of the credibility of evidence.

Theoretical framework

Defining the assessment of the credibility of evidence

We draw on Driver et al. (2000)'s work to provide an operational definition of evaluating the credibility of evidence. The credibility of different sources of information can be defined as the consideration of the grounds for confidence through the use of interrogatory tools, such as the critical asking of "reports about the origin of evidence, whether the evidence is simply correlational or whether there is a plausible theoretical mechanism, whether the results are reproducible, whether they are contested, or about the authority of the scientific source" (p. 301). Put simply, the researchers referred to two main parameters that are important for the assessment of the credibility of evidence: the source of the evidence and the methodology of the construction of the evidence (Driver et al., 2000). Learners need to examine the source of the evidence and think about questions such as: Is there evident bias or not? Was a piece of evidence peer-reviewed? Who is the author of the evidence? What is the author's agenda/background? What was the source of funding for producing each piece of evidence? As far as the methodology is concerned, learners need to think about the following: Does the evidence refer to a comparison of two different groups? Is there any control of variables? Were the results replicated?

Students' difficulties in evaluating the credibility of evidence

There are several reports in the literature on studies of credibility and how to better support students in their credibility evaluation skills; however, only a small subset of those studies come from science education, with most of them focusing on topics such as internet searches and evaluating information online. Science education research has demonstrated students' difficulties in evaluating evidence to construct evidence-based explanations at the middle school level (Glassner, Weinstock & Neuman, 2005), high school level (Sandoval & Millwood, 2005; Dawson & Venville, 2009) and undergraduate level (Lippman, Amurao & Pellegrino, 2008). One of the most important and common difficulty students have in relation to evaluating evidence refers to an uncertainty as to what constitutes convincing evidence or valid evidence (Driver et al., 2000). Furthermore, previous research showed that one of the important reasons students faced difficulties in evaluating the credibility of evidence was the fact that they lacked the criteria they needed to be able to evaluate evidence (Wu & Hsieh, 2006).

At the middle school level, Pluta, Buckland, Chinn, Duncan and Duschl (2008) documented a widespread difficulty that the participants in their study, 724 US middle school students, faced with regard to the effective use of reasons and evidence. Students typically found it difficult to make judgments about the relative strength of evidence, and rather tended to treat all evidence as equally strong. They also had difficulty in understanding the need to provide justifications that were more elaborated than just mentioning the evidence that provided support. Seethaler and Linn (2004) examined 190 8th grade students' reasoning about tradeoffs in the context of a technology enhanced curriculum about genetically modified food and found that even though students were able to provide evidence both for and against their positions, they were less explicit about how they weighed those tradeoffs.

At the high school level, Kolsto (2001) interviewed 22 10th grade students on their views on the trustworthiness of knowledge claims, arguments and opinions on a socio-scientific issue. All students expressed problems and uncertainty when trying to sort out who to trust and what to believe. As a result, some claims were not questioned at all. Moreover, most students showed very little interest in empirical evidence underpinning knowledge claims and only a few expressed a positive interest in methodological aspects to assess the credibility of evidence presented to them.

Even at the college level, students sometimes had difficulty assessing the credibility of evidence in scientific arguments. Lippman et al. (2008) investigated 98 college students' understanding of evidence use in quality scientific arguments, by measuring students' ability to compare and analyze the quality of arguments and the use of evidence in essays. Their findings revealed that students inconsistently applied the epistemological criteria of empirical data serving as evidence in scientific arguments. Their insensitivity to blatantly inaccurate descriptions of evidence suggests students lack the deep understanding needed to assess the quality of evidence. They made reasonable, but not always ideal selections for the strong and weak evidence and focused on superficial aspects, such as ease of comprehension instead of the need for relevant empirical and disconfirming evidence.

It seems that training can support students in evaluating evidence. With respect to evaluating evidence, Sanchez, Wiley and Goldman (2006) reported that college students had a fragile understanding of the credibility of sources of evidence in internet sites and very few could verbalize it or use it to justify their evaluations of credibility. The researchers identified four key areas in which college students needed support: considering the source of the information, considering the evidence that was presented, thinking about how the evidence fit into an explanation of the phenomena and evaluating the information with respect to prior knowledge. In their study with 60 undergraduates who were randomly divided into two groups, one group of students received a short

training course on criteria that could be used to evaluate internet sites such as: who the author of the information was, how reliable a site was, how well the site explained the information, identifying relevant information about the author, such as motivation, examining whether the information was consistent with other reliable sources and whether it was based on scientific evidence. Results indicated that college students were able to learn to use critical evaluation skills during short-term training and were better able to identify relevant from irrelevant sources.

The present study examined students' difficulties in evaluating the credibility of evidence during a decision-making process. The following research questions were addressed: How do 11th grade students evaluate the credibility of different sources of data when trying to reach a decision on a complex socio-scientific issue? What evidence do they use and how do they use it? How can this process be scaffolded?

Methodology

Participants

Two 11th grade intact classes in two public high schools in Cyprus participated in this study. The first class consisted of twelve students who took the course "Humans and Health" as an elective. The students of the class, in this case average students (mean GPA=14.7/20), were randomly assigned in four heterogeneous groups of three students each. The second class served as a control group and consisted of 13 students who took the course "Human and Health" as an elective in a different school.

The enacting teacher was a Biology teacher with a Bachelor's degree in Biology and 17 years of teaching experience. She is one of the authors and also a member of the group which designed the learning environment. The first author was present in the class during all lessons of the enactment to provide support with regard to the use of the web-based learning environment.

The intervention: The Biotechnology and Genetically Modified plants learning environment

A problem-based learning environment on the topic of Genetically Modified Organisms (GMOs) was developed using the web-based platform of STOCHASMOS (Kyza & Constantinou, 2007). STOCHASMOS is a learning and teaching platform for supporting students' scientific reasoning through scientifically authentic investigations with an embedded authoring tool that can be used by teachers or instructional designers. The driving question that students were asked to answer was whether they would allow the cultivation of genetically modified (GM) plants in their country. Students were asked to make sense of real and diverse data sets mostly coming from scientific publications that were comparing GM and conventional plants from the perspective of the environment, economy and health. Data took multiple formats and included sources of low, average and high credibility. An example of high credibility data was the result of experimental research (e.g. research that entailed the comparison of an experimental and a control group and was published in a peer-reviewed scientific journal). An example of low credibility data was an opinion-based article, which reflected the views of the author, was not accompanied by data and did not substantiate claims.

The activity sequence consisted of eleven 90-minute lessons. The first three lessons consisted of hands-on experiments with DNA extraction, while for the next eight lessons students worked collaboratively through the STOCHASMOS learning environment of Biotechnology. A 40-min section in the fourth lesson was devoted to having students evaluate the credibility of two data sources in a context unrelated to Biotechnology. The aim of this activity was to reveal the criteria that students were intuitively using to evaluate credibility and to guide them into identifying the following criteria that were part of the scaffolding within STOCHASMOS: "author background", "type of publication", "funding" and "methodological criteria". In lessons 5 to 10, students evaluated data, captured evidence and went through the process of interpreting and explaining the data organized in the STOCHASMOS WorkSpace templates to develop evidence-based explanations about whether they would allow the cultivation of GM plants in their country. Students were prompted to evaluate the credibility of evidence throughout their investigation. The lessons concluded with the groups' in-class presentations of their work.

Scaffolding through a reflective web-based inquiry environment

For the purposes of this research study, to "evaluate the credibility of evidence" means to assess the credibility and value of data that has been determined to be relevant for an inquiry-based investigation. Scaffolding was both human-provided and technology-supported. The scaffolding provided by the teacher included, among others, explicit instruction on how evidence and explanations are evaluated, based on given criteria. The scaffolding provided by peers included peer-evaluation of explanations, incorporated through the sharing work capability of the web-based platform used and the use of the chat-tool.

Technology-supported scaffolding was provided through the STOCHASMOS platform. The learning environment included the use of templates, which structured the students' task in creating evidence-based explanations for the impact of GM plants on economy, environment and health. The templates included passive prompts encouraging students to identify the source of the evidence and rate its credibility using a scale of 1 (low credibility) to 5 (high credibility). Drawing from the work of Driver et al. (2000), the main criteria for the evaluation of evidence in the context of this study were twofold and referred to: a) the source of the evidence and b) the methodology of the construction of the evidence. The source of the evidence referred to: author's background, type of publication and funding. Methodology was limited to having students examine whether the evidence referred to a comparison of two different groups, e.g. GM plants and conventional plants in this case.

Data collection and Analysis

When designing this case study we sought to collect diverse data that would maximize triangulation and increase the validity of the conclusions. Thus we collected both outcome and process data, at the individual students' level as well as at the collaborative, group level. Data sources included pre- and post-tests on students' conceptual understanding of Biotechnology concepts and the credibility of evidence and all groups' videotaped discussions throughout the enactment. The tests that were administered to the enacting group were also administered to the control group twice over a period of three months, which was the approximate duration of the enactment, to control for the testing effect. The control group students did not take any Biotechnology related lessons.

The conceptual understanding individual test consisted of ten multiple choice questions and seven open-ended questions on topics related to Biotechnology. A scoring rubric was developed for scoring students' answers by the first two authors. Interrater reliability was computed ($r=0.79$). A test examining individual students' evaluation of the credibility of sources was administered prior and after the enactment. The test consisted of two open-ended tasks. The first task asked students to evaluate the credibility of a given source. The rationale behind this task was to assess the criteria that students were intuitively using to evaluate credibility. In the second task, students had to compare the credibility of two sources to decide which cell phone model to buy. The rationale behind this task was to assess students' skill in differentiating between a credible (results of a study that used random sampling, control of variables, comparison of two groups, and was published in an independent magazine) from a non-credible publication (advertisement publication indicating source bias as it came from a company that sold the product and had economic interest). In both tasks students had to justify their answer. A scoring rubric was developed for scoring students' answers by the first two authors. Interrater reliability was computed ($r=0.83$). Students' responses were also examined qualitatively, using phenomenographic analysis methods to extract categories from students' answers.

All lessons during the enactment were videotaped. Two of the four groups of students, Group 1 (mean GPA=14.9/20) and Group 3 (mean GPA=16.2/20), were selected for in-depth analysis. The first researcher viewed all videotaped lessons twice to isolate meaningful episodes that related to issues of assessing the credibility of evidence; a list of relevant keywords (e.g. evidence, credibility, believable, source, author, bias, results, comparison, type of publication, funding, graph, internet, scientist, journal, and the rating that students used to evaluate credibility) was identified a priori to help guide this process. A total of 18 episodes were identified for Group 1, ranging from 1 to 10 minutes each (total= 61 minutes) and 14 episodes for Group 3, ranging from 1 to 4 minutes each (total=29 minutes). Three additional episodes referred to the teacher talking about credibility in whole class discussion activities. The episodes were transcribed in the local language. They were then analyzed with respect to the four criteria that students were asked to evaluate: author background, funding, type of publication and comparison between two groups, to identify students' difficulties and describe the development of students' thinking with regard to the credibility of evidence over time. We also analyzed the data to examine whether discussions were scaffolded and by whom or what (e.g. took place while students were working within the STOCHASMOS environment or took place after prompting by the teacher).

Results

Effectiveness of intervention

To assess students' learning gains, we compared their pre- and post-tests using the Wilcoxon signed ranks non-parametric test because of the small sample size. The analysis indicated a statistically significant difference $z(12)=-2.91$, $p<.01$, and an effect size of 0.59. The enacting students' performance increased from a mean of 16.58 (SD=5.02) in the pre-test to a mean of 24.17 (SD=6.13) in the post-test. The maximum possible score of the test was 38. The control group students' performance increased from a mean of 13.5 (SD=6.13) in the pre-test to a mean of 14.5 (SD=5.23) in the post-test. However, the analysis using the Wilcoxon signed ranks non-parametric test did not indicate a statistically significant difference for the control group students $z(12)=-1.27$, $p>.05$.

Students' evaluation of the credibility of sources pre- and post-enactment

<u>Results from the analysis of individual students</u>
The students' performance increased from a mean of 4.5 (SD=2.24) in the pre-test to a mean of 8.42 (SD=3.40) in the post-test (maximum possible score=12). To assess students' gains with regard to their ability to evaluate the credibility of sources, we compared their pre- and post-tests using the Wilcoxon signed ranks non-parametric test. The analysis indicated a statistically significant difference, $z(12)=-3.09$, $p<.01$, and an effect size of 0.63. The control group students' performance increased from a mean of 4.75 (SD=2.01) in the pre-test to a mean of 5.20 (SD=3.36) in the post-test (maximum possible score=12). The Wilcoxon signed ranks non-parametric test analysis did not indicate a statistically significant difference for the control group students $z(12)=-1.14$, $p>.05$.

How did students' understanding of the credibility criteria change? In addition to the quantitative analysis of students' skills in evaluating the credibility of sources, a phenomenographic analysis of their answers was also conducted. As a result, students' answers were categorized in levels which are presented in Tables 1 and 2. The analysis of the first task, which asked students to evaluate whether a publication was credible or not, showed that students' answers shifted from opinion-based answers (based on merely copying information from the text), to more sophisticated answers (which identified one or more criteria for evaluating credibility). Table 1 shows the number and percentage of students' answers in each level for the pre-test and post-test (n=12), as well as illustrative examples of the scoring.

Table 1:Task 1 "How did students evaluate the credibility of a given source?"

Category	Illustrative examples	Pre-test	Post-test
Level 1: Opinion-based/copying information from text	"I have the opinion that a person can only fight addiction with her own will (…) Therefore, I don't believe this publication".	3 (25%)	0
Level 2: Source provided details or numbers or results or statistics	"It (the research study) is not convincing because it does not mention details (…) On the other hand it mentions statistical results".	9 (75%)	0
Level 3: Based on one criterion (author, funding, source type, or methodology)	"The research study was conducted by the Acupuncture Center and it was published by the same center. Therefore they had interest in having positive results for advertisement purposes".	0	8 (67%)
Level 4: Based on at least two of the criteria	"The duration of the research study was adequate, the participants were men and women of various ages, the publication was from a university, the author was a doctor, a general practitioner. I don't think Dr. X. had any financial interest, because it's his duty to help people so I believe in this research study".	0	4 (33%)

The analysis of the pre-test showed that most students (75%) initially thought that if a source provided details and numbers, or results and statistics it meant that it was credible (Table 1). The rest of the students (25%) provided an answer that relied on either their own opinion or verbatim information from the text. However, none of the answers on the post-test fell into the pre-test categories. Students based their answers on the application of one (67%), or two and more credibility criteria (33%). This suggests that by the end of the intervention students had appropriated these criteria.

The second task asked students to choose which one of two given publications was the most credible. Table 2 shows the number and percentage of students' answers in each level for the pre-test and post-test for task 2 (n=12), as well as illustrative examples of the scoring in each level.

Table 2: Task 2 "How did students compare the credibility of two given sources?"

Category	Illustrative examples	Pre-test	Post-test
Level 1: No justification or opinion-based or copying information from text	"I think that the most convincing is the second publication because it seems more compelling".	4 (33%)	1 (8%)
Level 2: Source provided results	"It described the advantages of the cell phone in detail and provided specific results"	1 (8%)	0
Level 3: Identified "source bias" only	It was "an independent research study (…) published by an independent company".	1 (8%)	4 (33%)

| Level 4: Identified "comparison" only | "It presents a cell phone and compares it to another cell phone". | 5 (42%) | 0 |
| Level 5: Identified both "source bias" and "comparison" | "The publication was conducted by an independent company (…) and there was a comparison of two cell phones and the participants of the survey were chosen randomly". | 1 (8%) | 7 (58%) |

In the pre-test several students identified the fact that a comparison was conducted (42%), one student was able to indentify the bias of the source (8%), and another student identified both criteria (8%). In contrast, in the post-test, the vast majority of students (91%) correctly identified either "source bias" or both criteria.

Students' collaborative shift from opinion-based explanations to criteria-based explanations

An analysis of the 40-min section in the fourth lesson, which was devoted to having students evaluate the credibility of two data sources in a context unrelated to Biotechnology, was conducted to examine the criteria that students were intuitively using to evaluate credibility prior to working in STOCHASMOS. This showed that students were capable of identifying the "author background" criterion unassisted, but needed the teachers' support to identify other criteria for credibility, such as "type of publication", "funding" and "methodological criteria".

An analysis of the total of 35 episodes revealed that the vast majority of students' discussions on credibility of evidence (24 episodes) were prompted and scaffolded as they took place while students were working with the STOCHASMOS templates. Only two episodes took place spontaneously. The teacher prompted students to consider the evaluation of the credibility of their sources in two of the episodes, while another two episodes referred to teacher-initiated whole-class discussions. In the remaining five episodes students attempted to verify their sources. The analysis of students' videotaped discussions revealed several themes and illustrated the shift of the students of the two selected groups (n=6) from level 2 in the pre-test to levels 3 and 4 in the post-test.

Students became critical of and questioned the credibility of their sources

Baseline pre-test data showed that students sometimes even accepted advertisements as credible pieces of data and did not question the source of information. During the investigation, students became very critical of the sources they were basing their decision on. When Group1 students compared the expenses and income of GM to conventional maize they found the source non-credible. They expressed their concern that they were basing their conclusions about the effect of GM plants on economy on evidence that they did not find credible and they were puzzled by this fact. The group had some insecurity about their work and asked the teacher whether it was acceptable to make some observations and interpretations of data and reach some conclusions but then cite and evaluate their source only to find that they did not consider it a credible one. The same issue was troubling for students of Group 3.

Throughout the STOCHASMOS enactment and through their interaction with the learning materials and with their peers, students formed opinions about source credibility. These two groups' discussions about credibility focused on the need to evaluate all data sources and students gradually developed a critical stance toward the sources they were examining during their inquiry, which was evident in lessons 4 to 11.

Students showed distrust to sources of average credibility and were sensitive to source bias of "funding" and "type of publication"

As expected, sources of average credibility stimulated the most discussion among students during their evaluation. Students correctly showed distrust to those sources and were troubled as to what they should believe and what not. During the fourth lesson, Group 1 focused on the impact of GM maize on economy and examined financial data. The discussion showed the distrust of this group to the source, mainly because the study had been conducted by a financial advisor in a company that provided advice to biotechnology companies and it was published on the website of the company, receiving funding by the same company. In general, students showed distrust in anything that was published on the internet. Student names are pseudonyms.

Group 1: Lesson 4: min 2:24-11:00

Angelina: *"...to evaluate the credibility of the source. "Author background", Brook is the financial advisor of PG economics, which provides advice to biotechnology companies. I don't believe him because everything is done by the company. And they make biotechnology products. And in the comparisons, it says here that the GM is less than the conventional (talking about the expenses of cultivation) and he has the company for biotechnology so (...)"*
...I am not convinced. It's all for their benefit!"

> Kate: Now let's look at funding. Since it was funded by the same company it's not credible. (Should we rate it with) One (1) or two (2)?
>
> Angelina: One (1)!
>
> Kate: It's not convincing at all (...) because they published it on the website of their own company.

Similarly, in their attempt to evaluate the credibility of the same source, Group 3 students also identified funding as a potential problem and had the same strong idea that anything that was published on the internet was necessarily non-credible. However, it seemed that students gradually overcame that misconception. During lesson 8, Group 1 evaluated two sources on the impact of GM maize on people's health. Even though initially the group tended to believe that whatever was published on the internet was automatically non-credible, in this case they correctly identified that even though the publication was online, the source citation indicated that it was actually a journal publication.

Group 1: Lesson 8

> Rania: Wait. Yes, but (reading the title of the source), issue 29, p.26-28. Why would it then say publication on the internet? (laughing) Here I go again with my questions!
>
> Kate: Wait. Here it's not convincing because it was on the internet.
>
> Rania: Why does it say publication of the journal?
>
> Aggela: Ok, they took a part and then put it on the internet.
>
> Kate: Yes. So it **is** convincing (her emphasis).

Students easily identified journal publications as high credibility sources

Both Groups 1 and 3 had no difficulty evaluating a source that appeared in a peer-reviewed scientific journal, received funding from independent research organizations, was conducted by independent researchers or professors and included reported methodology that included a comparison of GM and conventional plants as a source of high credibility. Both groups rated all criteria with the highest possible score for credibility, which was 5 out of 5.

Students had difficulty in applying methodological criteria

Students faced some difficulty in understanding what the methodological criterion of "the existence of a comparison between two groups" meant and needed support for its application. Students did not seem to understand that the criterion "existence of a comparison between two groups" referred to having an experimental and a control group, even after four lessons of applying the credibility criteria during the investigation.

Discussion

The evidence suggests that the intervention was successful at increasing these students' conceptual understanding of Biotechnology-related concepts and at increasing their skills in evaluating the credibility of sources. Through their interaction with the learning environment, their teacher and their peers, students became critical of the sources of evidence. They paid attention to citing their sources and evaluating them when they were prompted to do so through the STOCHASMOS templates. They also paid attention to referring to their sources and commenting on their credibility during their in-class presentations of their work. Contrary to previous research, which documented middle school level U.S. students' difficulties in the effective use of reasons and evidence and stated their tendency to treat all evidence as equally strong (Pluta el al., 2008), the students in the present study had no difficulty in making judgments about the relative strength of evidence. This was evident by the finding that they correctly identified sources of low, average and high credibility. Students could also differentiate between weaker and stronger sources with regard to their credibility, as was evident by the analysis of their post-tests. This finding suggests that the intervention was successful with regard to supporting students' skills in evaluating the credibility of sources.

The fact that the majority of students' discussions regarding credibility criteria did not happen spontaneously but were, rather, a result of scaffolding showed students' need for support. The results of this study showed that a reflective web-based inquiry environment that prompted students to evaluate the credibility of their sources, combined with the use of given credibility criteria seemed to have provided an effective scaffolding mechanism toward the goal of making students develop a critical stance towards evidence. An epistemological shift was evident as students realized that there were two levels in evaluating the credibility of evidence: first an interpretation of data was necessary to make sense of the data that qualified as evidence and then as a second step, the credibility of the evidence could be assessed. They learned to question the evidence depending on the source it came from and at the same time they were puzzled and had reflective discussions about what to believe and what not, which would not have happened had they not been asked to engage in the task of evaluating the credibility of evidence they included in their investigation.

Implications for practice

This study demonstrated the importance, necessity and effectiveness of a reflective, web-based inquiry environment as a scaffolding mechanism to support students' evaluation of the credibility of evidence. An inquiry environment that is designed to support students in their credibility evaluation skills should include sources of data of varying degrees of credibility (e.g. low, average and high credibility) to provide students with opportunities to use critical judgment to accept and reject evidence. It was also shown that the design of interventions to foster students' credibility evaluation skills can be based on the use of specific criteria. Criteria can be either provided a priori or can be derived by students. Depending on the age, grade, and achievement level of the students, credibility criteria can be either simple and focus on two major parameters, such as "source" and "methodology" or can be advanced and sophisticated and cover a greater number of aspects and parameters of credibility. The level of complexity has to be decided in advance as it directly affects the amount of information that students need to have access to in order to be able to make an informed decision regarding how believable each credibility parameter is for them. Students in this study had difficulty in focusing on methodological issues of the studies they evaluated. It is important to address methodological criteria, which are among the most important ones, in relation to the evaluation of the credibility of evidence in future studies.

References

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically Authentic Inquiry in Schools: A theoretical framework for evaluating inquiry tasks. *Science Education, 86*, 175-218.

Dawson, V. & Venville, G. J. (2009). High-school Students' Informal Reasoning and Argumentation about Biotechnology: An indicator of scientific literacy? *International Journal of Science Education* 31 (11), 1421–1445.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the Norms of Scientific Argumentation in Classrooms. *Science Education*, 84(3), 287-312.

Gieryn, T. F. (1999). *Cultural Boundaries of Science: Credibility on the Line* (First ed.). Chicago: University Of Chicago Press.

Glassner, A., Weinstock, M., & Neuman, Y. (2005). Pupils' evaluation and generation of evidence and explanation in argumentation. *British Journal of Educational Psychology* (75), 105-118.

Kolsto, S. D. (2001). "To trust or not to trust,…"- pupils' ways of judging information encountered in a socio-scientific issue. *International Journal of Science Education,* 23 (9), p. 877-901.

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

Lippman, J. P., Amurao, F. K., & Pellegrino, J. W. (2008). Undergraduate Cognitive Psychology Students' Evaluations of Scientific Arguments in a Contrasting-Essays Assignment. Paper presented at the International Perspectives in the Learning Sciences: Cre8ing a Learning World, Utrecht, The Netherlands.

National Research Council. (1996). *National science education standards.* Washington, DC: National Academy Press.

Pluta, W. J., Buckland, L. A., Chinn, A. C., Duncan, R. G., & Duschl, R. A. (2008). Learning to Evaluate Scientific Models. Paper presented at the International Perspectives in the Learning Sciences: Cre8ing a Learning World, Utrecht, The Netherlands.

Sanchez, C. A., Wiley, J., & Goldman, S. R. (2006). Teaching students to evaluate source reliability during internet research tasks. Paper presented at the ICLS.

Sandoval, W. A., & Millwood, K. A. (2005). The Quality of Students' Use of Evidence in Written Scientific Explanations. *Cognition and Instruction*, 23(1), 23-55.

Seethaler, S., & Linn, M. (2004). Genetically modified food in perspective: an inquiry-based curriculum to help middle school students make sense of tradeoffs. *International Journal of Science Education,* 26(14), 1765-1785.

Wu, H. K., & Hsieh, C. E. (2006). Developing Sixth Graders' Inquiry Skills to Construct Explanations in Inquiry-based Learning Environments. *International Journal of Science Education,* 28(11), 1289-1313.

Yang, F. Y. (2004). Exploring high school students' use of theory and evidence in an everyday context: the role of scientific thinking in environmental science decision-making. *International Journal of Science Education*, 26(11), 1345-1364.

Acknowledgments