International Journal of Environmental & Science Education, 2015, **10**(6), 865-889

Cooperative Learning about Nature of Science with a Case from the History of Science

Balz Wolfensberger
University of Zurich, SWITZERLAND
Claudia Canella
University Hospital Zurich, SWITZERLAND

•*Received 13 March 2015* •*Revised 27 August 2015* •*Accepted 27 August 2015*

This paper reports a predominantly qualitative classroom study on cooperative learning about nature of science (NOS) using a case from the history of science. The purpose of the research was to gain insight into how students worked with the historical case study during cooperative group work, how students and teachers assessed the teaching unit, and in what ways students' ideas about selected aspects of NOS changed as a result of the teaching unit. In cooperation with two biology teachers, a four-lesson teaching unit about NOS and the early research on Archaeopteryx was developed, field-tested, modified, and tested again. Altogether, five classes of 10th and 11th grade students from two Swiss schools participated. Data were collected by videotaping group work, interviews with student groups and teachers, questionnaires, and pre- and post-tests about NOS conceptions. Results show that group work was mostly of good quality, both with regard to students' cooperation and understanding of the case study. Second, both the topic and the instructional design of the unit were judged very positively. Third, students showed more informed views on the selected target NOS aspects after the teaching unit. The paper ends with conclusions regarding teaching and learning about NOS, cooperative learning and questions for future research.

Keywords: nature of science, history of science, Archaeopteryx, classroom-based research, Cooperative Learning

INTRODUCTION

The overarching goal of science education is to promote *scientific literacy* (e.g. Kolstø, 2001; Laugksch, 2000; Millar & Osborne, 1998). Being scientifically literate implies not only having knowledge of the content and methods of science – that is, the laws, models, theories, concepts, techniques and procedures used by scientists – but also knowledge about the *nature of science* (e.g. Abd-El-Khalick, Bell, & Lederman, 1998; Bell, 2006; Laugksch, 2000). In other words, besides science content knowledge, scientific literacy encompasses knowledge of scientific inquiry and the nature of science (NOS). While scientific inquiry pertains to the processes 'by which scientific knowledge is developed', the nature of science refers 'to the epistemological underpinnings of the activities of science' and hence to the

Correspondence: Claudia Canella,
University Hospital Zurich, Institute for Complementary and Integrative Medicine
Sonneggstr. 6 CH-8091 Zürich, SWITZERLAND
E-mail: claudia.canella@uzh.ch
doi: 10.12973/ijese.2015.281a

ISSN: 1306-3065

'unavoidable characteristics' of scientific knowledge (Lederman, 2006, pp. 308f.).

Given that it is a vital component of scientific literacy, it is not surprising that a deeper understanding of NOS has been advocated as an objective of science education for decades. This holds especially true for the English speaking world and, more recently, also for Asian countries, South Africa, and Brazil (see e.g. Guerra, Braga, & Reis, 2012; Lederman, 2007; Tao, 2003). In Switzerland, by contrast, knowledge of NOS does not yet figure among the important educational outcomes for upper secondary education (see Schweizerische Konferenz der kantonalen Erziehungsdirektoren, 1994, pp. 105–115). While placing some emphasis on students' insights into the interplay between science, technology, society and nature, the national framework curriculum for upper secondary schools neither mentions NOS explicitly, nor does it implicitly refer to more than two of its aspects (though those it references are important): The historicity, and hence tentative nature of scientific knowledge, and the fact that there are questions science cannot answer.

One approach to teaching and learning about NOS is the exploration and interpretation of cases from the *history of science* (see Abd-El-Khalick, 2013). In the last two decades, a considerable number of instructional materials incorporating history of science (HOS) have been developed for science teaching. Moreover, guidelines for the design of historical cases to be used in teaching about NOS have been proposed (see Allchin, 2012; Clough, 2011; Guerra et al., 2012; Höttecke, Henke, & Riess, 2012; McComas, 2008; Schaake, 2011). On the other hand, empirical and especially qualitative classroom studies on implementation, learning processes, and the effects of using the history of science in teaching about NOS are not yet abundant. This is particularly lacking when it comes to cooperative learning in small groups.

The study reported here combined research and development. In cooperation with two experienced biology teachers and five upper secondary classes, we developed and trialled a teaching unit about the early research on Archaeopteryx, adopting a research design in which two cycles of development and research built on each other and were staggered over time. The study aimed to assess the teaching unit not only in terms of changes in students' views on selected aspects of NOS (the theory-laden, socio-culturally embedded, and tentative nature of scientific knowledge), but also in terms of how students worked cooperatively with the historical case, and how students and teachers experienced the teaching unit.

LITERATURE REVIEW AND RESEARCH QUESTIONS

In the following sections, we review the research literature on teaching and learning about nature of science and cooperative learning.

Teaching and learning about nature of science

Scholars in the fields of history, philosophy and sociology of science are quick to disagree on a specific definition of 'the' nature of science. However, for purposes of teaching NOS in schools, it seems that an acceptable degree of consensus about the most relevant aspects of NOS has emerged among science education researchers (e.g. Abd-El-Khalick, 2006; Clough & Olson, 2008; Lederman, Abd-El-Khalick, Bell, & Schwartz, 2002; Lederman, 2006; Osborne, Collins, Ratcliffe, Millar, & Duschl, 2003).

The NOS construct that is currently most influential in science education and which guided the research reported in the present paper was proposed by a research group led by Norman G. Lederman (e.g. Abd-El-Khalick et al., 1998; Lederman, 2006, 2007). According to their understanding, scientific knowledge: (1) is *tentative* (subject to change); (2) *empirically-based* (based on and/or derived at least partially from observations of the natural world); (3) *theory-laden* (involving individual or group interpretation); (4) *necessarily involves human inference, imagination, and creativity* (involves the invention of explanations); and (5) is *socially and culturally embedded* (influenced by the society/culture in which science is practised). The concept includes two additional important aspects: (6) the *distinction between observations and inferences*; and (7) the *distinction between scientific theories and laws* (Lederman, 2006, pp. 304; 2007, pp. 833f.). At the most general level, an understanding of NOS thus entails an understanding of its tentative nature. Accordingly, tenets (2) through (5) support the first tenet that scientific knowledge is subject to change and revision (Bell, 2006, pp. 430f.). However, there are much more complex interrelations between the NOS aspects (see Abd-El-Khalick & Lederman, 2000, p. 1064).

With regard to teaching approaches that aim to enhance an understanding of NOS, a considerable body of research shows that students and teachers do not automatically learn about the epistemological underpinnings of scientific knowledge simply by 'doing science'. Rather, an *explicit reflective approach* seems to be the most promising way of teaching NOS (see e.g. Bell, 2006; Clough, 1997; Lederman, 2006; Schwartz & Crawford, 2006; Smith & Scharmann, 2008). In this approach, instruction specifically and explicitly focuses on various aspects of NOS. It is important to note that the term 'explicit' does not refer to didactic or direct instruction. In fact, the term has curricular implications in that it 'entails the inclusion of specific NOS learning outcomes in any instructional sequence aimed at developing learners' NOS understandings'. In contrast, the term 'reflexive' has instructional implications, in that it calls for 'structured opportunities designed to help learners examine their science learning experiences from within an epistemological framework' (Abd-El-Khalick, 2013, p. 2091). Put differently, whatever NOS concept is adopted, its tenets 'might provide guidance for curriculum development, teaching, and assessment' (Clough & Olson, 2008, p. 143) but must not be misinterpreted as declarative knowledge to be transmitted rather than investigated in the classroom. Against this background, it is advisable to transform tenets into questions in order to explicitly challenge students' NOS views (Clough & Olson, 2008). Instructional approaches consistent with an explicit reflective approach are to be student-centred, active and constructivist (see e.g. Hofheinz, 2008; Lederman, 2006; Smith & Scharmann, 2008). This is further illustrated by calls for the inclusion of collaborative group work (e.g. Shipman, 2006), and for establishing a context in which students are encouraged to engage in reflexive, collaborative discourse on NOS aspects (e.g. Bartholomew, Osborne, & Ratcliffe, 2004; Nussbaum, 2008; Tao, 2003).

As a result of his review of the relevant literature, Lederman (2007, p. 869) concludes, inter alia, that: (a) neither students nor teachers typically possess 'informed' conceptions of NOS; (b) even when teachers hold 'informed' conceptions of NOS, these views do not automatically and necessarily translate into corresponding classroom practice; (c) more 'adequate' conceptions of NOS can be learned, the most effective approach thereby being explicit and reflective teaching as opposed to an implicit approach of merely 'doing science'.

Given the focus and design of our study, we will limit ourselves in the following to research literature that addresses the inclusion of cases from the history of science in teaching and learning about NOS. The call for incorporating the history of science (HOS) into science teaching has a tradition which can be traced back to the nineteenth century (for an overview, see Matthews 2012). Since the late 1980s, research and development in teaching about NOS with cases from the history of science have received even more interest (Teixeira, Greca, & Freire, 2012, p. 772).

One often-cited study by Abd-El-Khalick and Lederman (2000) found that history of science courses 'had only minimal influence on students' NOS views' (p. 1085). The authors conclude that teaching NOS with HOS might be more effective if it adopts an explicit approach and a conceptual change perspective that starts with eliciting the NOS views currently held by students and then invites them to test these views against specific historical examples (pp. 1088f.). Likewise, as a result of a study on students' sense-making during peer discussions about science stories, Tao (2003) comes to the conclusion that teachers should actively scaffold students' understanding, e.g. by conducting whole-class discussions following the peer collaboration activities, in order to prevent students from confirming and reinforcing their inadequate views on NOS (p. 168). Leach, Hind, and Ryder (2003) designed and trialled two single-lesson interventions using historical cases. The intervention did not prove effective in providing the majority of students with a more sophisticated understanding of NOS (p. 841). The authors interpret this finding in the light of the interventions' brevity. Moreover, they point out that many teachers, though judging the lessons very positively, were reluctant to explicitly tackle inappropriate ideas offered by their students during whole class discussions because they feared that this might inhibit contributions (p. 840).

By contrast, there are a number of recent studies reporting encouraging effects of instructional units that incorporate HOS into teaching about NOS. For example, in a qualitative action research study, Irwin (2000) shows that students who were taught subject matter whilst simultaneously learning about the historical development of atomic theory, clearly displayed less naïve-realist views about scientific theories following the teaching. Lin and Chen (2002) conducted a quasi-experimental study with pre-service chemistry teachers. Over one semester, students read and discussed materials that presented scientists' original debates and experiments. The experimental group showed a better understanding than the control group of the role of creativity in science, the theory-based nature of observations, and the functions of scientific theories (pp. 780–782). The authors interpret the results of their study as a consequence of, inter alia, the use of a 'student-centred instructional method of teaching', comprising small-group discussions and other cooperative learning activities (pp. 786f.). Rudge and Howe (2009) trialled an intervention that invited students to 'think along the lines that past scientists did as an exercise in thinking like scientists' (p. 566). After the eight-lesson intervention with 24 pre-intern elementary and middle school teachers, a marked change in students' understandings of four NOS aspects was found. Likewise, Paraskevopoulou and Koliopoulos (2011) found significant improvement in students' understanding of several NOS aspects after a five-lesson teaching intervention in which students learned about a historical scientific dispute by reading four short stories and answering accompanying questions focusing on different NOS aspects.

Cooperative learning and small-group discussions

In pedagogical theory and educational research, *cooperative learning* has been considered an important teaching approach for decades. Its development drew on the theoretical foundations of the socio-cognitive and socio-constructivist views of learning, largely grounded in the work of, e.g. Piaget, Vygotsky, and Dewey. In the 1980s the term 'cooperative learning' became prevalent in reference to 'small-group procedures' in the classroom (Davidson & Worsham, 1992, p. xiv) and to indicate a shift 'from a traditional classroom to one that is interactive and, further, to one that is cooperative' (Hertz-Lazarowitz & Miller, 1992, p. 5). Cooperative learning thus is characterised by, inter alia, student-to-student interaction in small groups of 2–6 students, a previously defined task or learning activity suitable for group work, the use of defined grouping procedures, shared leadership within groups, individual responsibility and accountability, and perspective-taking (Davidson & Worsham, 1992, p. xiii). Following Sharan (1990), the primary aims of cooperative learning are 'the improvement of student academic achievement and promotion of high-level thinking as well as positive interpersonal and inter-group relations among students in school' (p. 285).

In science education, small-group discussion work has become increasingly important as a teaching method (Oliveira & Sadler, 2008). Several factors – many of them related to the emergence of the concept of scientific literacy – have contributed to the growing interest in this approach. One of the most significant of these factors is the spread of constructivist views of learning. Another is the goal of preparing young people to participate as citizens in political decision-making on scientific matters, and yet another is the focus on the role of argument in understanding science (see Bennett, Lubben, Hogarth, & Campbell, 2005; Erduran & Jiménez-Aleixandre 2008; Osborne, Erduran, & Simon, 2004). With respect to the role of argument, an awareness of the social context of knowledge construction is especially important when it comes to discussions about the nature of science. It is argued that through cooperative learning, students can have an experience of how knowledge is constructed and negotiated, and therefore also grasp the process by which *scientific* knowledge is produced and tested (Erduran, Simon, & Osborne, 2004; Jiménez-Aleixandre, Bugallo Rodríguez, & Duschl, 2000; Sadler 2006; Zohar & Nemet 2002). In addition, students interacting with their peers might become more critical of the notion of 'absolute' scientific knowledge (see Wells, Chang, & Maher 1990, p. 98).

Positive effects of cooperative learning are reported across a broad spectrum: academic achievement, intergroup relations, acceptance of mainstreamed academically handicapped students, self-esteem, self-confidence, proacademic peer norms, locus of control, time on task and classroom behaviour, liking classmates, cooperation, altruism, and the ability to take another's perspective (Davidson & Worsham, 1992, p. xiv; Slavin, 1995, pp. 49–70). Stressing the importance of dialectic exchange within cooperative learning groups, Johnson and Johnson (1992) point out that the 'interpersonal exchange within cooperative learning groups, and especially the intellectual challenge resulting from conflict among ideas and conclusions (i.e. controversy), promotes critical thinking, higher level reasoning, and metacognitive thought' (p. 121). The authors conclude that 'meaning is formulated through the process of conveying it. It is while students are orally summarizing, explaining, and elaborating that they organize and systematize cognitively the concepts and information they are discussing' (pp. 122–125). Such findings are extended by research on social and cognitive processes of peer-group sense-making discussions in science classrooms, which suggests that having students work in cooperative groups does not necessarily lead to higher levels of reasoning and conceptual understanding even when students work on open-ended scientific inquiry tasks.

In a qualitative study, Hogan (1999) examined students' socio-cognitive roles during cooperative sense-making. Over a period of 12 weeks, eight small groups of 8th graders were observed while working on a prolonged task that required them to build mental models of the nature of matter, to use their models to explain phenomena, and to subsequently revise them. Analysis of students' group discourse resulted in the reconstruction of eight naturally emerging socio-cognitive roles played by the students during the mental model-building activity. While four of these roles ('promoter of reflection and regulation', 'contributor of content knowledge', 'creative model builder', and 'mediator of social interactions and ideas') were conducive to groups' reasoning and understanding, the four other roles ('promoter of distraction', 'promoter of acrimony', 'promoter of simple task completion', and 'reticent participant') hindered groups' reasoning processes (p. 877). In another study of the same instructional unit, Hogan, Nastasi, and Pressley (1999) examined interaction patterns and collaborative reasoning in four small groups, both with and without teacher guidance. The results showed that the four peer groups differed markedly in terms of the amount of discussion devoted to the nature of matter. This finding could not be explained by important differences in motivation among the groups but rather by the groups' varied ability to engage in 'productive dialogue'. Such dialogue evolved in groups where students presented 'provocative ideas', and were 'able and willing to ask for clarifications, and then interpreting and building on peers' ideas'. In addition, these groups' interactions were characterised by 'a consonant tone marked by overt acceptance of one another's ideas' (p. 424). When working alone, students tended to advance the discussion by sharing new ideas rather than by asking one another questions. By contrast, when a teacher was present, conceptual development was stimulated by his or her questions. Overall, the quality of reasoning – defined by features such as generation, elaboration, justification, and synthesis of ideas – was found to be higher in peer groups than in teacher-guided groups, with the only exception of explanations that emerged at a higher level when a teacher was present. Moreover, teacher guidance was found important to groups' progression when there was 'confusion or lack of synergy among their members' (p. 425).

Likewise, Oliveira and Sadler (2008) show that a teacher, when prompting students to expand and clarify their thinking, can make an important contribution to a group's conceptual understanding. The analysis of three small groups of elementary student teachers who collaboratively observed and explained the burning of a candle under a jar yielded marked differences in the groups' elaborative interactions and their ways of dealing with conceptual conflict – and consequently their levels of conceptual convergence. The authors explain these differences in terms of groups' diverging interpretations of the task ('intellectual endeavour' vs. 'worksheet completion') but also in light of socio-cultural variables like social status related to ethnic background. The results show that conceptual conflict seems to be important in stimulating the co-construction of conceptual understanding. However, as Oliveira & Sadler (2008, p. 655) conclude, in addition to 'having a friendly and positive social atmosphere' that allows conceptual conflict to emerge in a fruitful way, participants in cooperative group work also 'need to be willing to construct more elaborate explanations by combining and using ideas proposed by different group members, and to disagree with and challenge each other.'

Conclusions and research questions

Recent studies, though not yielding unequivocal findings, provide evidence for the idea that incorporating the history of science into science teaching can help students develop a more adequate understanding of NOS, especially if an explicit and reflective approach is adopted. Differences between research results may be accounted for due to differing conceptualisations of NOS, research designs, age groups, teaching approaches, content domains and so on. However, most of these studies have confined themselves to quasi-experimental and/or pre- and post-test designs exclusively aimed at investigating the effect of a specific HOS teaching intervention on students' NOS conceptions.

Group learning is a well-established field of research in science education. Research on cooperative group work has been related to students' understanding of science concepts, their explanations of scientific phenomena, or the resolution of socio-scientific issues. By contrast, there is a dearth of studies that investigate students' cooperative meaning-making about nature of science with a case from the history of science, as well as their experiences with learning within this context. The present study attempts to address this gap by asking the following research questions:

1. What is the content and process of students' small-group discussions about questions related to the historical case and the selected NOS aspects?
2. How do students and teachers experience and judge teaching and learning with this instructional unit?
3. In what ways and to what extent do students' views about the selected NOS aspects change as a result of the instruction?

DESIGN OF THE PROJECT

This predominantly qualitative research project was designed as a classroom study in which two cycles of design, evaluation and re-design were staggered over time and built on each other. In the first cycle of the project, we developed an initial version of the four-lesson teaching unit in cooperation with the teachers who then trialled it with their classes. On the basis of the findings from the subsequent field trials, we modified the teaching unit, again involving the teachers in the process. The modified unit was then trialled with the same teachers but different classes in the second cycle of field trials. The purpose of the present research was, within the theoretical background described above, to generate knowledge that is grounded in specific experiences in a real-life setting, and that can be useful for the field of practice under study (see Edelson, 2002, p.117f.). This should not imply that findings and explanations resulting from this type of research cannot be generalised at all. But rather, they are 'generalizable to theoretical propositions' and structurally similar cases as opposed to 'populations or universes' (Yin, 2003, p.10).

Participants

Two experienced biology teachers, Richard and Daniel[i], and 68 upper secondary students from five classes were involved in the project. The teachers voluntarily participated in the study during their regular class time. Both had an intrinsic interest in the history and philosophy of science as well as in teaching and learning about NOS but had not systematically taught this topic before. In Switzerland, high school students choose a so-called 'profile' after successful completion of grade 8 – that is for the remaining four years of their high school career. In addition to learning a range of subjects which continue to be taught in all profiles (e.g. mathematics; first, second and third languages; sciences; history, etc.), students will place an emphasis on a field of special interest. Of the five classes involved in this project, four were from the 'maths & science' profile (50 students, grade 11, 17–19 years old), and one from the 'art & design' profile (18 students, grade 10, 16–18 years old). Class size varied between six and 18. Three classes (38 students) were involved in the first cycle, and two classes (30 students) in the second cycle of the project.

Over the course of the project, the teachers and the research team met for four half-day workshops. In the *first workshop,* the teachers were introduced to the NOS concept, findings from studies on teaching and learning about NOS, and the design of the research study. We discussed our initial suggestions regarding a suitable historical case and potentially promising lesson designs with the teachers. Incorporating the outcomes of these discussions, the research team designed a preliminary version of the teaching unit comprising reading materials, student worksheets, and lesson plans. All materials were sent to the teachers for critical review, which formed the main subject of discussion in the *second workshop.* As a result of this discussion, the researchers revised the teaching unit so that it would be ready for field trials. The *third workshop* took place after our analysis of the data gathered in the first cycle of implementation and testing. The purpose of this meeting was to discuss the findings and the conclusions regarding subsequent revisions to be made to the teaching unit, with a view to the second cycle of field trials. Finally, the *fourth workshop* served as a venue to present and critically discuss with the teachers the research findings from both cycles of implementation and testing. We also sought the teachers' assessment of our hypotheses with regard to further revision of the teaching unit.

The teaching unit

As the literature reviewed above suggests, the teaching unit should follow a student-centred, active-constructivist framework, use a concrete example – in this project, a case from the history of science – and be explicit and reflective in addressing NOS.

Learning material

The core element of the learning material was a 20-page dossier containing a description of early years of research on the first Archaeopteryx fossil found in Germany in 1861. Choosing the case of Archaeopteryx seemed appropriate for at least two reasons. Firstly, it is especially illustrative of the tentative, theory-laden, and socio-culturally embedded nature of scientific knowledge. Secondly, Archaeopteryx, one of the most famous fossils ever found, can be easily fitted into the curriculum.

The historical case itself was presented in the form of a science story (Clough, 2011, p. 704), i.e. a chronological sequence of events involving protagonists – in this case, nineteenth-century scientists Johann Andreas Wagner, Richard Owen, and Thomas Henry Huxley –, and following a plot that relates these events. The story was told through narrative text and the protagonists' own words, but also included photographs, anatomical sketches, drawings and charts. Students' individual reading of the dossier as well as the collaborative group work was guided by questions to be answered in written form. Moreover, the dossier included basic information about Earth history, the process of fossilisation, and the current state of research on Archaeopteryx and the origin of birds. A final chapter introduced the students to the concept of NOS.

Structure and teaching methods

Table 1. Overview of the teaching unit

Lesson	Contents
1	Introduction: Science content, NOS
2	Reading dossier Preparing for group work
3	Small-group discussion
4	Whole-class discussion

Basically, both teachers followed the same structure of the four-lesson teaching unit over the two cycles: Firstly, the students were introduced to the science content relevant to the understanding of the Archaeopteryx fossil such as its place in Earth history, extant flora and fauna at the time, the process of fossilisation etc. Secondly, students read the dossier individually and then, as part of the reading assignment, answered questions that aimed at securing factual knowledge about the science content, e.g. 'give a brief description of flora and fauna in Solnhofen at the time of Archaeopteryx', and the science story, such as 'what did Huxley criticise about Owen's description of the fossil'.

Thirdly, in small group work, students discussed open-ended questions which invited them to make connections between the story and their ideas about scientific knowledge. On the one hand, they were asked to compare the scientific approach of the three protagonists, for example: 'how could Wagner, Owen and Huxley come to different conclusions even though they investigated the same fossil?'. On the other hand, they were guided to think about the progression of the scientific endeavour, e.g. 'how did scientific knowledge change in this case? And in general?' Finally, this small-group work was followed by a whole-class discussion that aimed at discussing the groups' findings as well as at drawing conclusions about the selected NOS aspects from the Archaeopteryx case.

Data collection and analysis

Data were collected firstly by *videotaping students'* collaborative small-group work and the subsequent whole-class discussions by using four cameras and eight microphones at the same time. In addition, the researchers also took field notes during the classroom activities. Overall, 11 student groups were recorded, to provide adequate variety. Secondly, *semi-structured interviews* with the groups and the teachers were conducted and videotaped. The purpose of the interviews was to elicit both student and teacher experiences with the teaching unit, and the students' interest both in the historical case and NOS. Moreover, the interview data also allowed for triangulation of our interpretation of student group work (see Miles & Huberman, 1994, pp. 266f.). Thirdly, student feedback on the learning materials was obtained by using a *short questionnaire* with open-ended questions. Fourthly, *open-ended pre- and post-test questionnaires* were administered in order to gain insight into students' ideas about NOS before and after working with the teaching unit. Pre- and post-tests were taken during class hours, the pre-test shortly before, and the post-test six weeks after the intervention. The questionnaire was an abridged version of the VNOS-C test (Lederman et al., 2002, p. 502), focusing on the target NOS aspects in question, and applied in the German translation developed and tested by Hofheinz (2008, p. 176f.).

All in all, the empirical data gathered in the project comprised 12 hours of classroom videos, 11 hours of interviews with students and teachers, 134 NOS questionnaires (67 pre-test, 67 post-test; return rate = 99%), and 62 feedback questionnaires (return rate = 91%). Video recordings were transcribed using the software 'Transana'.

In line with the research questions of this study, we chose a predominantly qualitative approach inspired by Naturalistic Inquiry (Lincoln & Guba, 1985). In the case of the pre- and post-tests, qualitative analysis was complemented by descriptive statistics. The specific procedures of analysis varied slightly according to the data: Transcripts of student group work were coded with existing and emergent categories. The most relevant of these form the italicised headings of section "Students' Small-Group Work". By contrast, interview transcripts and written feedback were coded with emergent categories, while pre- and post-test questionnaires were coded with existing categories; the most relevant of these are indicated in italics in the sections "Students' and Teachers' Experiences with the Teaching Unit" and "Changes in Students' Views of NOS". Regardless of these differences, qualitative analysis of all data sets always involved several iterations of close interpretive readings, including extensive writing of memos, aimed at reconstructing key themes and patterns in the data.

To enhance the trustworthiness of the research, the following procedures were employed (see Lincoln & Guba, 1985): Firstly, investigator triangulation was used to make the interpretation of data more credible. For each set of data, the analysis performed by one researcher was reviewed by the other. Divergent interpretations were revised collaboratively. As regards the pre- and post-test questionnaires, independent coding by two researchers could not be applied due to time constraints. However, the questionnaires were coded twice by the same researcher within a period of three months, with the resulting intra-rater reliability being >.80. Secondly, member checking with the participating teachers was applied during the third and the fourth workshops (see the subsection 'Participants' above). Thirdly, all raw data, methods of data collection and analysis, interpretive memos, working hypotheses, and provisional results were documented so as to provide an audit trail.

RESULTS

The main results of the study are reported in the following three subsections. The first focuses on students' small-group work. The second addresses student and teacher experiences with and assessment of the teaching unit. The final section presents findings on the students' NOS conceptions that emerged from the pre- and post-tests.

Students' small-group work

For this article, four groups (G2, G5, G7 consisting of four members, G9 of three), two of the first cycle, two of the second, are chosen out of the sample of 11 groups of which data were collected. The selected groups are representative for all groups under study here, and are especially illustrative in their way of discussing the historical case study of the Archaeopteryx-fossil and the related NOS-aspects as well as in their interaction during the group discussion.

Contents of the discussions regarding NOS

In both cycles of field trials, the majority of the student groups could be described as being well prepared for the group work. That is, the students remembered the important protagonists and events of the Archaeopteryx case study, or were quick to find the relevant information in the script. Most groups substantiated their answers by referring to information contained in the script. Within the group, they gathered together arguments and aspects, interchangeably verifying them. The groups did not have difficulty finding answers to the questions, but their arguments often remained implicit or unfinished. To give one example: Asked to compare the different proceedings of the early protagonists of Archaeopteryx research, Fabio of Group 9 (Fabio, Samuel, Tim) said with regard to Richard Owen: 'his science was based quite a bit on the church' (G9, Turn 65). His colleagues did not seem to have difficulty understanding this answer, which most probably referred to how Owen's world-view was inspired by biblical ideas and that his work as a scientist was influenced likewise. However, the meaning of Fabio's statement remains implicit and the question arises as to whether the student has concrete ideas about the ways in which beliefs or the adherence to a school of thought might influence scientists' work. In this context, it should also be mentioned that, during group work, students rarely drew conclusions from the science case study in terms of expressing more abstract, conceptual ideas about what constitutes science and scientific knowledge. However, students' ideas about NOS can be discerned in the group discussions, albeit on a scarcely explicit level. Particularly remarkable in this respect is that most groups oscillated in their view between an empiricist-realist and a constructivist view of science, which also led to inconsistencies in their ideas. A case in point is the discussion among G7 (Andrea, Anna, Luis, Sara) about whether every scientific theory is a matter of opinion, or if it is more likely that science gradually approximates the truth about natural phenomena. Sara initially defines goal and methods of science according to her opinion:

> (...) I mean, science theoretically should find out things and prove them.
> I think that is what they are here for. (G7, T 266)

However, she also adds that 'discoveries and inventions' are equally influenced by their 'discoverer' (G7, T 268–269). Andrea thinks that 'everything is based on facts' that are proven by the scientist using an experimental approach (G7, T 297). Anna tries to find a way out of the dilemma between the claim that science proves everything by experiments and the notion that scientific outcomes are influenced by the scientists' world-views:

> To be able to develop a theory one needs an opinion as a starting point
> which can be proved afterwards. (G7, T 293)

In a nutshell, the group comes to the conclusion that scientific knowledge is both a matter of opinion and truth: As a starting point, a scientist first needs an opinion in order to develop a theory, which then has to be proven by conducting appropriate experiments. Such a theory corresponds with objective knowledge and leads to true knowledge about natural phenomena. During this discussion, the students seemed to equate proving with being objective, without making this explicit. We found that the compromise this group came to in answering the question is also typical for most of the other groups.

In the following, we present in more detail the students' ideas about the theory-laden, socially and culturally embedded and tentative nature of scientific knowledge. However, it is important to keep in mind that the students themselves were not requested to discuss along the lines of explicitly-stated NOS aspects, but rather to explore open-ended questions, from which the research team assumed they would invite students to extract general ideas on NOS while still referring to the science story. Looking back, it may not be very surprising that the students never explicitly mentioned any NOS-aspect on a meta-level, but stayed close to the 'Archaeopteryx-story'.

Theory-ladenness and social and cultural embeddedness of scientific knowledge

'Theory-ladenness and social and cultural embeddedness of scientific knowledge' were assigned most often as a code. Referring to this aspect, three types of student associations can be discerned from the group discussions. The first type refers to the scientists as being mainly *driven by their own interests*. For example, G9 describes this as striving for 'becoming famous', 'being the better scientist' or 'proving that one's own world-view is correct', and therefore interpreting data in favour of one's own private goals (G9, T 16, 38). Referring to the script, Samuel adds that Richard Owen bought the fossil only to make sure that no one else could analyse the original, and that he would be the only one deciding which part of the fossil would be publicly accessible (G9, T 25).

Secondly, the students acknowledge the theory-ladenness and the social and cultural embeddedness of scientific knowledge in that it is depending on one's own world-view. This association may reflect the fact that each protagonist of the Archaeopteryx story had a different theoretical background, namely creationism (Wagner), spontaneous creation (Owen), and Darwinism (Huxley). Referring to these different backgrounds, the students explained the protagonists' diverging interpretations of the Archaeopteryx-fossil. For example, Luis of G7 states that every researcher had his own 'wishful thinking' about the Archaeopteryx (G7, T 9), and Sara complements this by saying that this was the reason the researchers could not change their minds (G7, T 10). It is Sara who adds that the followers of creationism and spontaneous creation could not accept the Archaeopteryx as being a 'mix' [missing link]. Only Huxley, the Darwinist, was able 'to bring it all together' (T 12), which we took to mean Huxley's ability to interpret the fossil as a bird with reptile characteristics. Later in this group discussion, Anna refers to another case, namely that Owen interpreted the head of the Archaeopteryx as a head of a fish not belonging to the rest of the fossil; Sara explains that this was because Owen 'wasn't able to explain the teeth of the fossil' (T 133–134) in the light of his theoretical background and his preconceived opinion that Archaeopteryx must be a bird. Similarly, Samuel argues that Owen 'interpreted the facts differently' (G9, T 13, 15), but it was Fabio who completes the argument that the researchers foremost wanted to win the 'power struggle' between creationism and Darwinism (G9, T17, 18). Moreover, most of the groups mainly equated world-view with religious belief. Tonio of G5 (David, Ilan, Jonas, Tonio) exemplifies this view by stating that Owen's 'religious beliefs' did not allow him 'any correct scientific argumentation'. Quite on the contrary, Owen 'bent' the interpretation of the fossil so that it would fit 'his opinion', Tonio adds (G5, T 179). Another case in point is Group 2 (Alexandra, Camilla, Dorothea, Nathalie) who, similarly to Group 7, ascribe the diverging interpretations of the different researchers to their religious beliefs, but also explicitly judge creationists as being less 'open to new things' than Darwinists (G2, T 162). It should be noted, however, that being critical of the ideas of creationism or spontaneous creation did not lead students to belittle the work of the historical scientists from their contemporary vantage point. Moreover, in no instance did the students suggest that, in contrast to past science, today's scientific knowledge would not be informed by theoretical backgrounds and socio-cultural contexts.

Thirdly, the students associate the NOS aspect in question here with the notion that scientists have to build *on each other* if they want to verify or falsify scientific claims. Anna of G7 expresses it as follows:

> (…) They [the scientists] wanted to prove that the others were wrong. In doing so, first of all, they were forced to examine the others' assumptions. They then may have had to admit: 'Yes, ok, after all it's maybe true what he said'. (G7, T 232)

Tentative nature of science

Statements coded as pertaining to the tentative nature of science can be found particularly in the context of one question, which asked the students to discuss what could cause a change in scientific knowledge – with regard to the Archaeopteryx case as well as to science in general. Merging their answers to the two sub-questions, the students came to the conclusion that scientific knowledge changes firstly as a result of 'new findings' (G5, T 396), such as when Huxley found out that the 'head of the fossil was twisted' (G5, T 398), or when 'new fossils' are found; that is, when 'scientific data are expanded' and thus allow scientists 'to say something more precise' (G5, T 400). Secondly, the students claimed that scientific knowledge changes in the light of 'new theories and hypotheses' 'on empirical material', as was the case with the Archaeopteryx fossil being interpreted on the basis of the evolution theory (G5 T 398, 400/ G2, T 361). Finally, 'technological progress' (G5, T 429 / G2, T 361) and 'funds' (G5, T 471–478) were mentioned as contributing to the change of scientific knowledge. While G2 added that 'cultural tendencies' (G2, T 379) also influence scientific knowledge, G5 discussed more specifically the influence of 'religious beliefs' on possible change of scientific knowledge. Although they did not reach full consensus within the group, they agreed that 'religious beliefs' may lead to a change of scientific knowledge in such a way that religious beliefs 'prevent' or 'retard' the 'progress of science' (G5, T 432–439).

Another argument of the students related to the tentative nature of science was that scientists build on each other's work in terms of 'falsifying' or 'verifying each' other's assumptions and conclusions, mainly by using experiments. We have already seen variations of this argument in the paragraphs on general content and on the theory-ladenness of scientific knowledge. Consequently, the idea that scientists build on each other's work is a very popular belief among the groups. G9 explicitly elaborates on this argument, discussing whether every scientific theory is a matter of opinion, or (more likely) if science gradually approximates the truth about natural phenomena. Fabio states (G9, T 125):

> I mean, simply, science progresses gradually and they once in a while falsify an old opinion, and classify it as false and develop a new one, which is better, I better say, which can be falsified less easily. And some day, one gets to the point where one can say, now that's right and everything else isn't. And as a result of this, I think that the sciences gradually progresses until they reach a point someday, where they can say, that's the way it is and nothing other than this.

Tim, by contrast, does not believe that science could 'ever reach this point', rather 'only comes close', because at some point there are no more 'counterexamples which can be falsified by experiments' (G9, T 126, 128).

Features of collaboration and interaction during group work

The discussions of cycle 1 can be characterised as assembling the answers of the individual group members and assimilating them into the group's records, whereas the discussions of cycle 2 were clearly more transformative, as the students exchanged their opinions about the issues in question and strove for mutual understanding. This can be explained by the fact that, in cycle 2, all questions to be discussed in the groups were open-ended. Moreover, the students received the questions only at the beginning of the group discussions while, in cycle 1, they had already prepared their answers individually before.

In both cycles of field trials, the students worked in a highly task-oriented manner, and spent most if not all of the time allowed for the collaborative group work on completing the assignments. This result could be explained, on the one hand, through the students' high levels of interest in the topic and the novel way of approaching it (both were expressed in the post-lesson interviews, see the following subsection). On the other hand, it could be put down to the presence of the research team with their video cameras, which may have prompted the students to commit themselves more seriously to the tasks than they would normally: For example, we observed a sudden digression from the task in one group when a member of the research team had to leave the room due to a technical problem. As the researcher went out of the room, the students immediately stopped working on their assignment and started to chat. Nevertheless, after a short time, they agreed that they should not digress from the task because the camera was still recording, and they went on with the discussion. The groups who showed less knowledge of the case study were also less concentrated in their work. 'Less concentrated', however, does not imply that the students neglected their assignment. Rather, they were less specific in their references to the case study and in their argument.

Normally, the students' statements were very short and often incomplete, and turn-taking was frequent. Sometimes students talked simultaneously. However, this does not imply that they did not listen or refer to each other. An illustrative example of this is given in the following sequence, taken from a Group 7 discussion about the scientific approach of the three protagonists of the Archaeopteryx story (square brackets denote simultaneous speaking, aborted sentences are marked by a slash):

> Andrea: [Yes, partially they] proceeded in the same way. They all have something with each other/ so, [they compared it with each other]/
> Anna: [Yes, they did it with other things, which they] knew already/
> Andrea: Yes, they compared it, so, with, I mean, [birds and/]
> Luis: [But, for one's own] they all had their own. (Anna: Mhm, Sara: Mhm).
> Anna: And he, so/ [they just were]/
> Andrea: [They just were/ so, simply proceeded in a similar way]
> Sara: [So yes and no?]
> Luis: [Yes].
> Andrea: [Yes].

Considering this spoken way of arguing, as researchers and teachers we might have to put into perspective the ideal that students' discourse about NOS should be 'argumentative'. According to our observations, it seemed less important than we had thought for the students to formulate their statements carefully, precisely and in whole sentences in order to gain an understanding of the case study and aspects of NOS.

The students treated each other in a friendly and respectful way, hearing each other out, relating to each other, involving each other in the discussion, and accepting different opinions without making fun of each other or devaluing diverging opinions. They sought to incorporate the ideas of all group members into the group's answers. Especially in the case of G2, the students sometimes went so far in trying to take account of all contributions that they even accepted incorrect answers without addressing the resulting inconsistencies. Furthermore, the students were eager to complete the assignment by gathering and assembling information from the script as efficiently as possible. In doing so, the students were consistently concentrated and tempered in their gestures, their facial expressions and their voice levels. Thus, the majority of the students worked well together in the team and also fulfilled their roles in a very natural and smooth way. However, most of the students did not consider it necessary to assume predetermined roles during group work, with the exception of appointing a minute-keeper.

Regarding the remarkably high task-orientation and concentration of the students during the group work, combined with their respectful interaction, we hypothesise that, as a result of the presence of the research team and their cameras, many of the students were to some degree 'doing school'. That is, they reproduce group work according to their ideas of how group work should look like ideally within the 'institution of school', and thus make cultural norms visible. This hypothesis is supported by statements from the interview with Group 9. They confirmed that they behaved more 'respectfully' and worked in a more 'disciplined' way than usual due to presence of the camera (G9, Interview, T 117). They added that, as a result, they became aware of how a successful group discussion proceeds and that this would help them doing better group work in the future:

> [...] this lasts, I think, that we rather, in future discussions without cameras, without supervision, that we can maintain the discipline to reach a better result.' (G9, Interview, T 118)

Students' and teachers' experiences with the teaching unit

In response to the open-ended question at the beginning of the interview ('how did you experience these past lessons?'), ten out of the 11 groups being interviewed said that this teaching unit was 'something (completely) different', 'something new', or a ‘welcome change from the routine’ of school. In three groups, students explicitly made the argument that novelty equals interest:

> Yeah, sure this [i.e. Archaeopteryx] ain’t a bad example, ‘cause I haven’t heard about it before. And that it’s half bird and half reptile was also new to me. [...] Well, that was something new. And, well, ‘new’ means ‘interesting’, mostly. (Group 10, Turn 47)

In talking about novelty and interest, the groups referred to the topic of the teaching unit, the instructional design, or both.

Overall, the *topic* of the teaching unit was judged ‘interesting’, sometimes ‘very interesting’ or even ‘exciting’ by the students. Despite some minor disparities between groups and their enjoyment of the different facets of the topic (i.e. Archaeopteryx as an animal, the historical case study of research on Archaeopteryx, and nature of science), the historical case study was well received throughout. Moreover, only one out of 11 groups explicitly said that they were ‘not very interested’ in finding out about the characteristics of scientific knowledge. They explained that this was due to the topic being too ‘theoretical’ (G8, T 42f., 59, 61, 65). Though, according to one member of the group, NOS-related questions became ‘a little more interesting’ after discussing them in the small group, the historical case study and the concept of NOS remained mostly unrelated until the end of the project (see G8, T 49, 53).

The *design* of the teaching unit was judged favourably by the majority of the groups. As emerged from the interviews, this may in part be attributed to the students’ highly positive attitudes towards *small-group work*, both in general and in looking back at this teaching unit in particular. The students stressed the following advantages of the small-group work: ‘being and discussing by themselves’; the individual student having ‘more opportunities to make a contribution’ to the discussion; ‘gathering a variety of ideas and opinions by complementing and correcting each other’, and ‘learning and retaining things better by having to verbalise them’ in the group. In other words, students, according to their assessment of the group work, experienced ‘involvement’ and ‘autonomy’ (e.g. by listening to and making sense of other students’ ideas, and working among peers without teacher intervention), which are seen as important features of social constructivist learning environments (McRobbie & Tobin, 1997). Moreover, students’ appreciation of ‘complementing’ and ‘correcting’ each other may indicate that a prerequisite of group cooperation has been met, as there was ‘resource interdependence’ in the groups. That is, group members needed each others’ resources in order to achieve the group’s goals (E. G. Cohen, 1994, p. 12).

As the small-group work was the highlight of the teaching unit, this might have produced a ‘halo effect’, that is, the positive impression of one of its components might have contributed to judging the teaching unit positively as a whole. Moreover, while all groups reported that pair- or small-group work was also used in ‘regular’ lessons, they agreed that the group work conducted in connection with this study was different from anything they had previously experienced. Thus, as seems to be the case with the judgement of the topic of this instructional unit, a ‘novelty effect’ might also have had a bearing on students’ positive assessment of the teaching method. However, one piece of critical feedback remained constant throughout both project cycles: All groups objected to the predetermined roles to be assumed during small-group work (i.e. time-keeper, moderator, keeper of the minutes, group spokesperson), claiming they were unnecessary, with the exception of the keeper of the minutes.

When asked whether their *ideas about NOS* changed as a result of the instructional unit, the students gave rather cautious assessments. Half of the groups said that their views regarding the selected NOS aspects had changed. Though the other half did not say the unit was useless with regard to better understanding NOS, they described the learning outcome not so much as a 'change', rather that their pre-existing NOS views were 'consolidated', 'specified' or 'enriched'. Besides these commonalities in students' self-assessments, one difference should also be noted: In cycle 2, students who reported a change in their NOS views were more specific in describing the nature of this change than those in cycle 1. One group was particularly remarkable as they not only described what they had learned but also explicitly contrasted their new insights with the rather sterile and objectivist view about science and scientists normally presented in school (G9, T 42–46). As one student put it:

> [...] there is this ideal: 'Scientists do this [i.e. research] for the world to know more', and the like. But you never hear in the classroom that they are as human as non-scientists (smiles), that they, too, are not just disinterested. One is concerned with the theories only, not with the actions and thoughts of the humans behind them. (G9, T 45)

Turning to the *teachers' reflections,* it was significant that both teachers were positively surprised by the quality of the cooperation among the students in the small groups and the outcome of the group work, as regards students' understanding of the Archaeopteryx case study.

In their reflections on the instructional design, both teachers considered the final lesson of the unit crucial, because it is at this point that students should progress from understanding a specific example (i.e. the history of science case) to developing more general, conceptual ideas about NOS. In both cycles of the project, the teachers were dissatisfied with this last lesson, albeit for different reasons. Richard said that he generally felt uncomfortable with the question-and-answer method of instruction; on the other hand, Daniel deemed it highly appropriate given that the aim was to reflect about NOS, but he criticised the lack of time, which was due to a need to discuss and secure the results of the group prior to talking about the nature of science on a more conceptual level.

Changes in students' views of NOS

Table 2. Descriptive statistics for Aspect-Specific Adequacy Quotients (AQ) and effect size values, by cycle of implementation and testing

		Pre-Test			Post-Test			Effect Size
Cycle	**NOS Aspect** [a]	***N***	***M***	***SD***	***N***	***M***	***SD***	***d*** [b]
1 [c]	Theory-ladenness / Embeddedness	37	.19	.29	37	.69	.30	1.72
	Tentativeness	37	.42	.48	37	.66	.44	.52
2 [d]	Theory-ladenness / Embeddedness	30	.33	.41	30	.72	.37	1.00
	Tentativeness	30	.43	.49	30	.50	.45	.15

[a] In the coding system adopted from Hofheinz (2008), target NOS aspects, "theory-ladenness", and "social-cultural embeddedness" were collapsed into one category.

[b] Cohen's d = $\frac{Mpost-Mpre}{SDpooled}$*, where* $SD_{pooled} = \sqrt{\left[\frac{(SDpre^2+SDpost^2)}{2}\right]}$*. J. Cohen (1988) suggests, that d values of ≥ .20, ≥ .50, and ≥ .80 represent small, medium, and large effect sizes.*

[c] Classes A, B, C.
[d] Classes D and E.

In coding the *pre- and post-test questionnaires,* we used the coding scheme developed and successfully tested with the German version of VNOS-C by Hofheinz (2008, p. 283f.) The answers were coded for target NOS aspects, thereby also distinguishing between 'rather naïve' and 'rather adequate' views. For the purposes of quantitative analysis, an 'adequacy quotient' (AQ) was computed by dividing the number of units coded as 'rather adequate' by the total number of coded units. Consequently, AQ values can vary between 0 and 1, with higher values denoting more adequate views of NOS. Quantitative analysis employed descriptive statistics and the computation of effect sizes. By contrast, no inferential statistics were applied, given that the data came from a real-life context with a multitude of uncontrollable variables, let alone that the classes under study were not random samples (see Shaver, 1993).

Table 2 shows the pre- and post-test AQ scores for the target NOS aspects focused on in this research project.

First, as regards students' views on the *theory-ladenness & social-cultural embeddedness* of scientific knowledge, the data for both cycles of the project reveal important gains between pre- and post-test adequacy quotient scores, high post-test AQ values, and – according to J. Cohen's (1988) rules for interpreting *d* values – large effect sizes. This substantial increase in adequacy may be explained by the fact that the *theory ladenness & social-cultural embeddedness* of scientific knowledge lay at the heart of our account of early Archaeopteryx research, and was also addressed explicitly by questions to be discussed during small-group work. While effect sizes are large in both cycles, there are differences between the two cycles. In contrast to post-test AQ values being very close to each other, the students in cycle 2 started from a higher mean pre-test score which consequently made the difference between pre- and post-test smaller. The higher AQ pre-test value can be ascribed to the singularly high score of one of the two classes that participated in cycle 2 (.41 compared to values between .14 and .28 for the other four classes). The available data do not allow a robust explanation for this score; the difference might as well be random. At least, the interviews with the students and the teacher of the class in question show that these students had not received any training with regard to NOS previous to our project which could explain the difference.

Second, a medium-sized improvement was found for views about the *tentativeness* of scientific knowledge, but only among the students involved in cycle 1. In cycle 2, by contrast, the improvement was negligible in terms of Cohen's *d*. Another feature with regard to this target NOS aspect were the comparatively high pre-test scores in both cycles of the research.[ii] When trying to explain the varying extent of pre- and post-test changes in the two cycles (effect sizes of .52 in cycle 1 compared to .15 in cycle 2), it should be pointed out that, in cycle 1, the tentativeness of scientific knowledge was much more explicitly addressed by one of the small-group discussion questions than in cycle 2. The reformulation of the groups' assignment in cycle 2 thus seems to have resulted in a considerably smaller change in students' views about the tentativeness of scientific knowledge. To explain the high pre-test levels of acceptance of the idea that scientific knowledge, including theories and laws, is subject to change, we hypothesise that students might link the notion of tentativeness primarily with the notion of available data (yet) being 'incomplete' and 'inconclusive' (see Oliveira, Akerson, Colak, Pongsanon, & Genel, 2012, p. 678).

In the remainder of this section, we will turn from an aspect-specific to a more integrative analysis of students' NOS views. Qualitative analysis of the data suggests that the changes between pre- and post-tests can also be described as an emerging coexistence of decreasing but by no means disappearing realist-naturalist views on the one hand, and constructivist-culturalist views of science and scientific knowledge on the other, where the latter became markedly more frequent after the intervention. Put differently, another distinctive feature of pre- and post-instruction change is that, on a more general level, students' views about NOS became more inconsistent.[iii]

One example out of many shall suffice to illustrate this inconsistency:

Answer to question 1 ('What, in your view, is science? What makes science, or a scientific discipline such as physics, biology, etc., different from other disciplines of inquiry, e.g. religion, philosophy?'):

> In the sciences, things are researched by use of experiments, tests etc. You can prove exactly how something is or was. You have something tangible.
>
> In other disciplines of inquiry like philosophy or theology, things are neither tangible nor provable. They rely only on ideas which are in one's mind. Every person has a different view.

Answer to question 3 ('How can two groups of scientists develop widely accepted but differing hypotheses about the extinction of the dinosaurs from the same set of data?'):

> Every person, and hence also every scientist, is interpreting each piece of information, each piece of evidence etc. differently. Thus, they might come to different conclusions. Moreover, scientists can be influenced by opinions and ideas of others.
>
> (Student 19, post-test)

Question 1 asked the students in a very open manner about what they think science is and what makes it different from other disciplines of inquiry like religion or philosophy. The answer of student 19 to this question reflects an empiricist, realist and naturalist view of scientific knowledge: Scientific knowledge is seen as 'exact', 'provable', 'tangible' – and, by implication, not relying on 'ideas', as is the case in other disciplines. Likewise, a number of other students' answers suggest that they believe scientific knowledge to be something that is simply 'received' by the scientists from their observations of nature – almost as if nature itself was granting insights directly to the scientist – when in fact it is something that must be constructed via interpretation of data. By contrast, the same student in response to question 3 claimed that 'every scientist is interpreting' data and acknowledged that, in doing so, he or she might be influenced by certain 'opinions and ideas'.

The question arises as to why this inconsistency does not seem to have led to a cognitive conflict in the student – and her colleagues who expressed similar views in the post-test. A plausible hypothesis might be that the students themselves did not perceive their views as inconsistent at all. This may in turn be explained by the specific content of the teaching unit and the different focus of the NOS test questions 1 and 3: While question 1 elicited ideas about 'science' *in general,* question 3, in asking why two groups of scientists could develop widely accepted but differing hypotheses about the extinction of the dinosaurs from the same set of data, addressed a *specific field* of scientific research, namely palaeontology. After the teaching unit, a great number of students admitted that palaeontology involves interpretation of data (i.e. question 3). This might not be surprising, given that the case study about the early research on Archaeopteryx employed in the classroom concerned palaeontology as well, and the conflicting interpretations of the fossil during these years were, to a great extent, due to researchers' differing theoretical and ideological backgrounds. However, the insight into the interpretive and theory-laden nature of scientific knowledge gained from the example of palaeontology hardly affected the more general, naïve realist and empiricist notions of science evoked by question 1. Concurring with Abd-El-Khalick (2006, pp. 417–419), we might argue that students' openness to the notion of scientific knowledge as theory-laden varies with different scientific disciplines under consideration, and that, from their point of view, palaeontology is not equally 'scientific' as, for example, physics.

In summary, the analysis showed that the students' views more often than not changed in the desired direction. However these views were often neither 'rather adequate' or 'rather naïve', but a sometimes complex intermixture of both. The most notable finding in this regard is the growing tension between views about scientific knowledge as being objective and true on the one hand and theory-laden and embedded in a socio-cultural context on the other: While post-instruction views about the former remained fairly stable, students became receptive to new, more adequate views about the latter. Consequently, the pre- and post-instruction changes cannot be described as the supplanting of consistent 'rather naïve' views with consistent 'rather adequate' ones. This finding resonates with the self-assessments of roughly 45% of the interviewed students.[iv] When asked whether and in what ways their understanding of scientific knowledge changed as a result of the instruction, they said that their views have been 'enriched' rather than 'changed'.

DISCUSSION AND CONCLUSIONS

There are limitations that should be considered when interpreting the results of this study. First, despite including a range of data sources, relatively few students and teachers from only two schools were involved in this project. More research would be needed to examine whether our findings can be transferred to a wider spectrum of upper secondary classes, e.g. in terms of students' study profiles, their previous knowledge about NOS and experience with the kind of group work used in this study (which both tended to be low for the students in our sample), urban vs. rural context of schools, etc.

Second, in examining students' conceptions of NOS, written tests and observations of group discussions, but not interviews, were used as data sources. Interviews would have provided a richer qualitative picture of students' conceptions and allowed for a more accurate interpretation of the quantitative results. In this study, however, the interviews served a different purpose, which was to gain insight into how the students experienced and judged the teaching unit.

With few exceptions (e.g. Tao 2003), previous classroom-based studies of teaching and learning about NOS have confined themselves to an analysis of participants' pre- and post-intervention NOS views. By contrast, the present study also focused on what happened in the classroom when students were working with a historical case study, and how teachers and students experienced and judged the instructional unit.

The field trials of the Archaeopteryx teaching unit with five upper secondary classes have yielded a number of encouraging results.

First of all, collaborative group work was of good quality in the majority of cases, both with regard to students' understanding of the science story and in terms of their high levels of time-on-task, issue-related communication and respectful interaction. The students themselves judged this teaching and learning method very positively. They felt not only challenged but also enabled to verbalise their individual chains of thought, express their opinion and, more importantly, to exchange their ideas and results among themselves without teacher intervention. Consequently, they reported that they had gained a richer, more complete understanding of the case study in question. Considering all these aspects, collaborative group work is a powerful method when it comes to making sense of a relatively long and complex case study. According to statements from the interviews, the fact that they were subject to research had a motivating effect on at least some of the groups. However, this is not to say that the quality of the group work should be played down as 'only a Hawthorne effect'. Rather, we agree with Brown's claim that a Hawthorne effect might be exactly what teachers and researchers aim for when engaging in cooperative classroom research (see Brown, 1992, pp. 163–167).

A second finding is the wide acceptance of the teaching unit among the students. As most of the interviews with the groups showed, this favourable assessment can in part be explained by the perceived novelty of both the topic and the instructional design – with the small-group work apparently being the highlight of the unit. In addition, and in contrast to the findings of McRobbie and Tobin (1997), students explicitly valued autonomy in learning. Thus, the perceived autonomy made possible by the small-group work may also have contributed to students' positive judgment of the teaching unit.

Thirdly, as regards the contents of students' NOS views emerging from both the small-group discussions and the pre- and post-tests, our findings are consistent with the results of previous studies, e.g. that views on the epistemological status of scientific knowledge are reflective of a tendency towards ontological realism, and that views on the procedures and methods of science tend to be naïve-empiricist. Also, we concur with Abd-El-Khalick's (2006) observation that NOS conceptions are often inconsistent, fragmented, and fluid. Beyond that, however, the analysis of the pre- and post-tests also allowed us to describe how views changed as a result of the instruction. On the target NOS aspects focused in this research – 'theory-ladenness & social-cultural embeddedness' and 'tentativeness' – students' views started to change in the desired direction. Qualitatively, this change is best described as follows: NOS views became more inconsistent as a result of the intervention; constructivist-culturalist notions increased markedly while realist-naturalist notions decreased, but by no means disappeared. From a conceptual change point of view, this finding should not be interpreted as a shortcoming but rather as an opportunity for explicitly addressing and discussing inconsistent ideas about NOS with students.

This being said, the group work observed in the present study should also be discussed in view of the issue of conflict. *Interpersonal conflicts* were practically non-existent in the groups, which may in part be explained by the fact that students were allowed to choose their own group members. In turn, the absence of interpersonal conflicts could explain the groups' high levels of time-on-task (see Hogan, 1999, p. 879). While minimising interpersonal conflict is conducive to collaboration, the presence of *cognitive* or *conceptual conflict* between group members – 'willingness to disagree and challenge each others' ideas' – facilitates a group's conceptual gains (Oliveira & Sadler, 2008, p. 651). In our study, however, instances of conceptual conflict among group members were observed only rarely. When trying to explain this finding, we should take into account that students' (and teachers') interactions 'are embedded within a task context' (Hogan et al, 1999, p. 426). Thus, and drawing on Oliveira and Sadler (2008, p. 653), the rare occurrence of conceptual conflict might reflect students' understanding of the task. More specifically, the majority of students in our project might have interpreted the groups' task in terms of completing a worksheet, rather than as a 'true' inquiry assignment that required them to develop a deeper understanding by challenging and elaborating each others' ideas as to what the historical case revealed about NOS. Group 2, described earlier in this paper, provides a particularly good illustration for students that seemed to place a higher value on completing a worksheet than identifying and resolving potentially conflicting views. Another possible explanation for groups' avoidance of conceptual conflict is that, in a social context that is not friendly and supportive, challenging a group member is risky (Oliveira & Sadler, 2008, p. 652). However, in view of the positive social atmosphere we observed, this explanation is not likely to account for our findings; still, some groups might have established shared social norms that inhibit challenging each others' views (see Hogan, 1999, p. 877).

There are a number of interacting factors that affect groups' discussion patterns (see Hogan, 1999; Oliveira & Sadler, 2008). At this point, however, we confine ourselves to a conclusion regarding the task of the group work. Arguably, our project tasks, despite being open-ended and pertaining to a problem that does not have one right answer (Cohen, 1994, p. 8), were not framed in a way that encouraged students to elaborate more clearly on one another's ideas. Therefore, it may prove fruitful to reframe the assignment as a problem to be solved by the group. For instance, the groups could be asked to come up with their own definition of scientific knowledge and the ways in which it is produced. In doing so, they would be required to support their claims with evidence from the historical case, and, possibly, to agree on additional features of science which might not be represented in the case.

To a certain extent, the historical case developed and trialled in this study was found to be suitable in enhancing students' understanding of the 'human side' of the scientific endeavour. Conversely, other HOS cases like the sickle-cell unit designed and tested by Rudge and Howe (2009) may be more effective in addressing the *empirical nature of science* or the *difference between observation and inference,* which, in the case of the sickle-cell story, can be attributed to the fact that students actively 'engage[d] in the sort of reasoning that led past scientists to reach insights about scientific phenomena' (p. 561) instead of 'only' analysing what past scientists did. Yet another case that could be powerful in teaching and learning about NOS is the story of Ignaz Semmelweis and his fight against puerperal fever. This case would allow students to engage in and reflect on hands-on inquiry activities (e.g. using Semmelweis' original data in order to build and test hypotheses) as well as to consider a rich, even dramatic account that illustrates the influence that social, cultural and political environments as well as existing theories have on the production and acceptance of scientific knowledge.

This being said, we remain doubtful whether there is a single HOS case that would address all aspects of the NOS construct equally well. What is more, explicit-reflective teaching about NOS with case studies from the history of science is only one of several approaches that can be effective in enhancing students' epistemological understanding of scientific knowledge production and NOS (see e.g. Abd-El-Khalick 2013). For instance, there is evidence that explicitly and reflectively addressing NOS within inquiry-based laboratory instruction can be effective, in terms of learners' views becoming more adequate (for a recent account, see Ozgelen, Yilmaz-Tuzun, & Hanuscin, 2013). Also, using non-lab inquiry activities related to science process skills (e.g. observing, inferring, or hypothesising) as an authentic context for nature of science teaching has yielded good results (e.g. Bell, Mulvey, & Maeng, 2012). We assume that a combination of approaches and a repeated engagement with NOS would prove most effective. To our knowledge, no research has yet been undertaken to systematically compare the relative effectiveness of different approaches or combinations thereof, nor are there long-term studies comparing the effect of students' repeated engagement with NOS over several years with the effect of onetime interventions.

Another question future research should attend to is how to 'develop teachers' valuing of NOS as an important instructional outcome' (Lederman, 2007, p. 872). There are studies describing promising ways to develop teachers' content and pedagogical content knowledge (PCK) and to increase their intention to actually teach NOS (e.g. Akerson & Abd-El-Khalick, 2003; Hanuscin, 2013; Schwartz & Lederman, 2002). Quite a different question, however, is how to integrate NOS into teacher education curricula. For the elementary- and middle school levels, Switzerland is just about to implement standards that include knowledge of inquiry and nature of science. By contrast, the national framework curriculum for secondary schools (which is not obligatory), does not list NOS among the educational objectives.[v] In light of this situation (and for the time being), we think it is appropriate to follow a more modest, bottom-up approach like promoting development and research cooperation between researchers/educators and intrinsically motivated pioneer teachers, as was done in this project. Such cooperation could be one way of raising wider interest in and knowledge about teaching NOS if this results in the provision of teaching materials and teaching strategies that have been developed and successfully tested in the real-life context of classroom teaching and learning.

Moreover, if we aim at raising teachers' valuing of research on teaching and learning about NOS and reducing the gap between theory and practice that often affects the adoption of curriculum innovations (Irwin, 2000, p. 13), future classroom-based research should not be confined to testing the effectiveness of an intervention in terms of pre- and post-test results while ignoring what is actually going on in the classroom and how students (and teachers) perceive and judge specific learning goals, teaching materials and instructional methods. The present study took a step in this direction.

ACKNOWLEDGMENTS

We cordially thank the teachers and students for their participation in this project. We would also like to extend our gratitude to Regula Kyburz-Graber for acting as a project supervisor, and to Jolanda Piniel who contributed significantly to the recruitment of the teachers and the development of the teaching unit. Last, but not least, thanks to the four anonymous reviewers for their constructive comments on an earlier version of this paper, and to Ciara Murray for editing the manuscript. This study was supported by the Swiss National Science Foundation (SNSF) under Grant No. 100014-124556.

REFERENCES

Abd-El-Khalick, F. (2006). Over and over and over again: College students' views of nature of science. In L. B. Flick, & N. G. Lederman (Eds.), *Scientific inquiry and nature of science. Implications for teaching, learning, and teacher education* (pp. 389–426). Dordrecht: Springer.

Abd-El-Khalick, F. (2013). Teaching *with* and *about* nature of science, and science teacher knowledge domains. *Science & Education, 22*(9), 2087–2107.

Abd-El-Khalick, F., Bell, R. L., & Lederman, N. G. (1998). The nature of science and instructional practice: Making the unnatural natural. *Science Education, 82*(4), 417–436.

Abd-El-Khalick, F., & Lederman, N. G. (2000). The influence of history of science courses on students' views of nature of science. *Journal of Research in Science Teaching, 37*(10), 1057–1095.

Akerson, V. L., & Abd-El-Khalick, F. (2003). Teaching elements of nature of science: A yearlong case study of a fourth-grade teacher. *Journal of Research in Science Teaching, 40*(10), 1025–1049.

Allchin, D. (2012). The Minnesota Case Study Collection: New historical inquiry case studies for nature of science education. *Science & Education, 21*, 1263–1281.

Bartholomew, H., Osborne, J., & Ratcliffe, M. (2004). Teaching students 'ideas-about-science': Five dimensions of effective practice. *Science Education, 88*(5), 655–682.

Bell, R. L. (2006). Perusing Pandora's box. Exploring the what, when, and how of nature of science instruction. In L. B. Flick, & N. G. Lederman (Eds.), *Scientific inquiry and nature of science. Implications for teaching, learning, and teacher education* (pp. 427–446). Dordrecht: Springer.

Bell, R. L., Mulvey, B. K., & Maeng, J. L. (2012). Beyond understanding: Process skills as a context for nature of science instruction. In M. S. Khine (Ed.), *Advances in nature of science research: Concepts and methodologies* (pp. 225–245). Dordrecht: Springer.

Bennett, J., Lubben, F., Hogarth, S., & Campbell, B. (2004). *A systematic review of the use of small-group discussions in science teaching with students aged 11-18, and their effects on students' understanding in science or attitude to science*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences, 2*(2), 141–178.

Clough, M. P. (1997). Strategies and activities for initiating and maintaining pressure on students' naive views concerning the nature of science. *Interchange, 28*(2–3), 191–204.

Clough, M. P. (2011). The story behind the science: Bringing science and scientists to life in post-secondary science education. *Science & Education, 20*, 701–717.

Clough, M. P., & Olson, J. K. (2008). Teaching and assessing the nature of science: An introduction. *Science & Education, 17*(2–3), 143–145.

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*(1), 1–35.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Davidson, N., & Worsham, T. (Eds.). (1992). *Enhancing thinking through cooperative learning*. New York: Teacher's College Press.

Edelson, D. C. (2002). Design-research: What we can learn when we engage in design. *The Journal of the Learning Sciences, 11*(1), 105–121.

Erduran, S., & Jiménez-Aleixandre, M. P. (Eds.). (2008). *Argumentation in science education. Perspectives from classroom-based research*. Dordrecht: Springer.

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.

Guerra, A., Braga, M., & Reis, J. C. (2012). History, philosophy, and science in a social perspective: A pedagogical project. *Science & Education, online first, DOI: 10.1007/s11191-012-9501-5*.

Hanuscin, D. L. (2013). Critical incidents in the development of pedagogical content knowledge for teaching the nature of science: A prospective elementary teacher's journey. *Journal of Science Teacher Education, online first; DOI: 10.1007/s10972-013-9341-4*.

Hertz-Lazarowitz, R., & Miller, N. (Eds.). (1992). *Interaction in cooperative groups. The theoretical anatomy of group learning*. Cambridge: Cambridge University Press.

Hofheinz, V. (2008). *Erwerb von Wissen über 'Nature of Science'. Eine Fallstudie zum Potenzial impliziter Aneignungsprozesse in geöffneten Lehr-Lern-Arrangements am Beispiel von Chemieunterricht*. Dissertation, Universität Siegen, Siegen.

Hogan, K. (1999). Sociocognitive roles in science group discourse. *International Journal of Science Education, 21*(8), 855–882.

Hogan, K., Nastasi, B. K., & Pressley, M. (1999). Discourse patterns and collaborative scientific reasoning in peer and teacher-guided discussions. *Cognition and Instruction, 17*(4), 379–432.

Höttecke, D., Henke, A., & Riess, F. (2012). Implementing history and philosophy in science teaching: Strategies, methods, results and experiences from the European HIPST project. *Science & Education, 21*, 1233–1261.

Irwin, A. R. (2000). Historical case studies: Teaching the nature of science in context. *Science Education, 84*(1), 5–26.

Jiménez-Aleixandre, M.-P., Bugallo Rodríguez, A., & Duschl, R. A. (2000). 'Doing the lesson' or 'doing science': Argument in high school genetics. *Science Education, 84,* 747–792.

Johnson, D. W., & Johnson, R. T. (1992). Encouraging thinking through constructive controversy. In N. Davidson & T. Worsham (Eds.), *Enhancing thinking through cooperative learning* (pp. 120–137). New York: Teacher's College Press.

Kolstø, S. D. (2001). Scientific literacy for citizenship: Tools for dealing with the science dimension of controversial socioscientific issues. *Science Education, 85*(3), 291–310.

Laugksch, R. (2000). Scientific literacy: A conceptual overview. *Science Education, 84*(1), 71–94.

Leach, J., Hind, A., & Ryder, J. (2003). Designing and evaluating short teaching interventions about the epistemology of science in high school classrooms. *Science Education, 87*(6), 831–848.

Lederman, N. G. (2006). Syntax of nature of science whithin inquiry and science instruction. In L. B. Flick, & N. G. Lederman (Eds.), *Scientific inquiry and nature of science. Implications for teaching, learning, and teacher education* (pp. 301–318). Dordrecht: Springer.

Lederman, N. G. (2007). Nature of science: Past, present, and future. In S. K. Abell, & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 831–879). Mahwah, NJ/London: Erlbaum.

Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., & Schwartz, R. S. (2002). Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Science Teaching, 39*(6), 497–521.

Lin, H.-S., & Chen, C.-C. (2002). Promoting preservice chemistry teachers' understanding about the nature of science through history. *Journal of Research in Science Teaching, 39*(9), 773–792.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills: Sage.

Matthews, M. R. (2012). Changing the focus: From nature of science (NOS) to features of science (FOS). In M. S. Khine (Ed.), *Advances in nature of science research* (pp. 3–26). Dordrecht etc.: Springer.

McComas, W. F. (2008). Seeking historical examples to illustrate key aspects of the nature of science. *Science & Education, 17*(2–3), 249–263.

McRobbie, C., & Tobin, K. (1997). A social constructivist perspective on learning environments. *International Journal of Science Education, 19*(2), 193–208.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis. An expanded sourcebook* (2. ed.). Thousand Oaks / London / New Delhi: Sage.

Millar, R., & Osborne, J. (Eds.). (1998). *Beyond 2000: Science education for the future*. London: School of Education, King's College.

Nussbaum, E. M. (2008). Collaborative discourse, argumentation, and learning: Preface and literature review. *Contemporary Educational Psychology, 33*, 345–359.

Oliveira, A. W., Akerson, V. L., Colak, H., Pongsanon, K., & Genel, A. (2012). The implicit communication of nature of science and epistemology during inquiry discussion. *Science Education, 96*(4), 652–684.

Oliveira, A. W., & Sadler, T. D. (2008). Interactive patterns and conceptual convergence during student collaborations in science. *Journal of Research in Science Teaching, 45*(5), 634–658.

Osborne, J., Collins, S., Ratcliffe, M., Millar, R., & Duschl, R. (2003). What 'ideas-about-science' should be taught in school science? A Delphi study of the expert community. *Journal of Research in Science Teaching, 40*(7), 692–720.

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

Ozgelen, S., Yilmaz-Tuzun, O., & Hanuscin, D. L. (2013). Exploring the development of preservice science teachers' views on the nature of science in inquiry-based laboratory instruction. *Research in Science Education, 43*, 1551–1570.

Paraskevopoulou, E., & Koliopoulos, D. (2011). Teaching the nature of science through the Millikan-Ehrenhaft dispute. *Science & Education, 20*(10), 943–960.

Rudge, D. W., & Howe, E. M. (2009). An explicit and reflective approach to the use of history to promote understanding of the nature of science. *Science & Education, 18*, 561–580.

Sadler, T. D. (2006). Promoting discourse and argumentation in science teacher education. *Journal of Science Teacher Education, 17*(4), 323–346.

Schaake, S. (2011). *Die Natur der Naturwissenschaften verstehen lernen: Historische, gesellschaftliche und kulturell relevante Stationen für den Chemieunterricht* (Reihe Studium und Forschung; Bd. 17). Kassel: kassel university press.

Schwartz, R. S., & Crawford, B. A. (2006). Authentic scientific inquiry as context for teaching nature of science: Identifying critical elements for success. In L. B. Flick, & N. G. Lederman (Eds.), *Scientific inquiry and nature of science. Implications for teaching, learning, and teacher education* (pp. 331–356). Dordrecht: Springer.

Schwartz, R. S., & Lederman, N. G. (2002). 'It's the nature of the beast': The influence of knowledge and intentions on learning and teaching nature of science. *Journal of Research in Science Teaching, 39*(3), 205–236.

Schweizerische Konferenz der kantonalen Erziehungsdirektoren (1994). *Rahmenlehrplan für die Maturitätsschulen.* Bern: EDK.

Sharan, S. (1990). Cooperative learning: A perspective on research and practice. In S. Sharan (Ed.), *Cooperative learning. Theory and research* (pp. 285–300). New York: Praeger.

Shaver, J. P. (1993). What statistical significance testing is, and what it is not. *The Journal of Experimental Education, 61*(4), 293–316.

Shipman, H. L. (2006). Inquiry learning in college classrooms: For the times, they are, a changing. In L. B. Flick, & N. G. Lederman (Eds.), *Scientific inquiry and nature of science. Implications for teaching, learning, and teacher education* (pp. 357–387). Dordrecht: Springer.

Slavin, R. E. (1995). *Cooperative learning. Theory, research, and practice* (2nd ed.). Boston, Mass.: Allyn and Bacon.

Smith, M. U., & Scharmann, L. (2008). A multi-year program developing an explicit reflective pedagogy for teaching pre-service teachers the nature of science by ostention. *Science & Education, 17*, 219–248.

Tao, P.-K. (2003). Eliciting and developing junior secondary students' understanding of the nature of science through a peer collaboration instruction in science stories *International Journal of Science Education, 25*(2), 147–171.

Teixeira, E. S., Greca, I. M., & Freire Jr, O. (2012). The history and philosophy of science in physics teaching: A research synthesis of didactic interventions. *Science & Education, 21*, 771–796.

Wells, G., Chang, G. L. M., & Maher, A. (1990). Creating classroom communities of literate thinkers. In S. Sharan (Ed.), *Cooperative learning. Theory and research* (pp. 95–121). New York: Praeger.

Yin, R. (2003). *Case study research: Design and methods* (3ed., Applied Social research Methods Series; 5). Thousand Oaks / London / New Delhi: Sage.

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.

---

[i] All participants' names have been changed.

[ii] In assessing scores for this NOS aspect, it should be considered that for a response to be coded as 'rather adequate' it was not sufficient to simply state *that* scientific knowledge is tentative (the latter was acknowledged by 88% of the students before, and by 96% after the instruction). Rather, the response also had to reflect an understanding of factors that may lie behind the change of scientific knowledge.

[iii] Here, the word 'inconsistent' does not denote an idea that is in complete disagreement with currently accepted 'adequate' conceptions, but rather a notion that incorporates both 'rather naïve' and 'rather adequate' facets. Other authors use the term 'transitional view' to label this phenomenon (see e.g. Ozgelen et al. 2013, p.1558).

[iv] It should be noted that only students who belonged to one of the videotaped groups were subsequently interviewed. In cycle 1 this applied to 17 (46%), in cycle 2 to 20 students (67%).

[v] This is not to suggest that the inclusion of NOS in the national curriculum would automatically result in it being widely taught in secondary schools and teacher education. Nevertheless, it would be an important argument in discussions about secondary teacher education reform.