International Journal of Science Education

ISSN: 0950-0693 (Print) 1464-5289 (Online) Journal homepage: http://www.tandfonline.com/loi/tsed20

A teaching–learning sequence on a socio-scientific issue: analysis and evaluation of its implementation in the classroom

Ángel Vázquez-Alonso, Abdiel Aponte, María-Antonia Manassero-Mas & Marisa Montesano

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION, 2016
VOL. 38, NO. 11, 1727–1746
http://dx.doi.org/10.1080/09500693.2016.1212289

Routledge
Taylor & Francis Group

A teaching–learning sequence on a socio-scientific issue: analysis and evaluation of its implementation in the classroom*

Ángel Vázquez-Alonso[a], Abdiel Aponte[b], María-Antonia Manassero-Mas[a] and Marisa Montesano[c,d]

[a]Center of Postgraduate Studies, University of the Balearic Islands, Palma, Spain; [b]University of Panamá, Panamá City, Panamá; [c]Department of Psychology, University of the Balearic Islands, Palma, Spain; [d]SENACYT, Panamá City, Panamá

ABSTRACT
This study examines the effectiveness of a teaching–learning sequence (TLS) to improve the understanding of the influences and interactions between a technology (mining) and society. The aim of the study is also to show the possibility of both teaching and assessing the most innovative issues and aspects of scientific competence and their impact on the understanding of the nature of science. The methodology used a quasi-experimental, pre–post-test design with a control group, with pre–post-test differences as the empirical indicators of improved understanding. Improvements were modest, as the empirical differences (pre–post and experimental–control group) were not large, but the experimental group scored more highly than the control group. The areas that showed improvement were identified. The paper includes the TLS itself and the standardized assessment tools that are functional and transferable to other researchers and teachers.

ARTICLE HISTORY
Received 9 July 2016
Accepted 9 July 2016

KEYWORDS
Scientific literacy; nature of science and technology; science–technology–society; teaching–learning sequences; assessment of improvement; quasi-experimental design

In the didactic literature, scientific literacy – or scientific culture – has two basic components (Hodson, 2009; Millar, 2006): (i) knowledge 'of' science, which includes traditional knowledge on the facts, concepts, principles and processes of science and (ii) knowledge 'about' science, which involves understanding how science and scientists build and validate knowledge. In science education, this last component incorporates interdisciplinary contents (adapted to the students' level) relating to the philosophy, history and sociology of science, and is referred to as 'nature of science'.

An important aspect of the nature of science is its relationship to technology. Although this relationship has changed greatly throughout history, it has become so close nowadays that experts have recognized as a new construct called 'technoscience'. When dealing with education and science, technology and society (STS) issues, this strong interaction between

CONTACT Ángel Vázquez-Alonso angel.vazquez@uib.es

*This paper first appeared in *Educ. Quim* (2014), 25(1), 190–202, a journal in Spanish. It appears by kind permission of the Editor, Professor Jose Chamizo. It was chosen because it illustrates the value of studies that use a standard procedure to address learning in a novel context.

science and technology (S&T) can be embodied, by analogy, in the integrated term of 'nature of science and technology'. This term will be used hereafter to describe how S&T interact with today's society (Tala, 2009). Nature of science and technology (herein after NoST) is a body of interdisciplinary metaknowledge on S&T, about what and how S&T interact in the world. The core content of NoST is the construction of scientific knowledge, which includes issues of epistemology – the philosophical principles underlying the validation of knowledge – as well as issues regarding STS relationships. These relationships include the internal sociology of science (focused on the social construction of knowledge, the scientific community and the work of scientists), S&T interaction (outlined in the previous paragraph) as well as the external sociology of S&T (the interactions between society and the scientific and technological system), which is the focus of this study (Bennàssar, Vázquez, Manassero, & García-Carmona, 2010).

As the NoST aspect is a basic component of scientific and technological literacy (Millar, 2006), specialists in the didactics of science consider that it is important enough to be included in education. Actual and effective teaching of this content in the classroom, however, represents a major innovation challenge to which, as for any educational innovation, there is typically a resistance, as shown by the non-institutionalization of NoST in the classroom and little attention given to it in teacher training (Matthews, 1994). In addition, NoST contents are not properly covered in textbooks, either because their development is left to teachers, or because textbooks provide an outdated and distorted view of science: positivist science. This conception of science does not agree with current proposals based on the philosophy, history and sociology of S&T, or with the complexity, interdisciplinary nature and novelty of NoST itself (McComas & Olson, 1998). Complexity is the internal factor that represents the greatest conceptual challenge to incorporating NoST into course contents suitable for teaching in the classroom. However, specialists now propose some ideas and issues that have become the subject of a broad consensus, no longer a controversy, hence reducing this complexity (Acevedo, Vázquez, Manassero, & Acevedo, 2007; Lederman, 2007; Matthews, 2012; McComas & Olson, 1998; Osborne, Collins, Ratcliffe, Millar, & Duschl, 2003; Vázquez-Alonso & Manassero-Mas, 2012).

The overwhelming conclusion of extensive research is that students do not have an adequate, informed and accurate understanding of NoST (Abd-El-Khalick & Lederman, 2000; García-Carmona, Vázquez, & Manassero, 2012; Lederman, 2007). The NoST ideas of students of various ages and from different countries are dominated by an absolutist/empiricist perspective of science. Another indicator that NoST is poorly understood is that scientific training has proved to be ineffective in promoting an adequate understanding of NoST. Indeed, science students do not obtain better scores when compared to humanities students, even when this comparison is made in the context of teacher training (Bennàssar et al., 2010; Vázquez & Manassero, 2008).

In recent years, these poor results have strongly driven research on NoST teaching and learning, involving students at different educational levels and from different backgrounds, to identify the factors that determine the effectiveness and improvement of NoST teaching in the classroom. Despite the complexity of the issue, a review of the literature allows us to identify two key conditions for teaching effectiveness in the classroom (Acevedo, 2009; Hodson, 2008; Lederman, 2007):

(i) The explicit teaching of NoST contents: intentionally and clearly addressing the specific aspects of NoST, which implies planning all curricular elements – objectives, contents, activities, methodology and assessment – consequently and explicitly applying these aspects in class. Making NoST explicit in the curriculum is an alternative to the 'implicitness hypothesis', which claims that it can be learned vicariously through content and activities in which the presence of NoST is indirect or implicit.
(ii) The development of reflective activities on NoST by the students: metacognitive activities (e.g. analysis, argumentation, discussion, debate, conclusion and exploration) should be developed by them in class to reflect on NoST. Critical reflection by students is an active alternative to listening to the teacher's reflection.

Teaching NoST is not intended to train philosophers or sociologists; it only pursues the modest goal of training scientifically literate persons, achievable by all (Matthews, 1998). This entails that the view of S&T that is taught should be at once authentic, critical, balanced between the various issues, adapted to the educational level of students, functional (that is, useful both in their daily lives and to better understand S&T theories) and, finally, based on proposals of specialists (Hodson, 2008). In a recent review, Deng and colleagues (2011) show that 88% of the studies using explicit approaches reported statistically significant or recognizable improvements in NoST understanding, while only 47% of those using implicit approaches did. In addition, three studies comparing implicit and explicit approaches also agreed in detecting positive changes in NoST understanding when the matter was taught explicitly, but no change when it was taught implicitly.

The most effective context for teaching NoST remains under discussion, as researched contexts have been many, varied and mixed: inquiry, history of S&T, socio-scientific issues, incorporation of NoST into course contents and absence of context (black-box activities of the type 'Guess what is in … ?'). The specific teaching contexts actually leading to innovative teaching of NoST are the history of S&T and issues of social interest that include technical and scientific content, where a public issue (e.g. mining, as in this study) is analyzed and debated by the students with the purpose of deepening understanding of the social and ethical implications of S&T, conflicting interests and the integration of social and epistemic norms (Hodson, 2009; Zeidler, Sadler, Simmons, & Howes, 2005). The study of Deng, Chen, Tsai, and Chai (2011) also shows that the most frequent reflection strategies involved in NoST teaching are discussions, whether dependent or independent of content (45%), reflection (26%) and argumentation (16%). These authors confirm that all studies using argumentation or reflection led to improvements in NoST understanding, and so did most of the studies using some kind of debate, but studies lacking a reflective activity produced no change. It thus seems that reflective activities are necessary to effectively improve NoST understanding.

Teaching NoST is a complex task, as the line of demarcation between right and wrong is blurred, has many nuances, is problematized (critical) and changes with time and the progress of knowledge, and because in NoST, questions are just as important as principles (Clough, 2007). This means that NoST should be taught in the actual context in which knowledge arises – including controversies and the logical and social competition between winning and losing ideas, with the important thing being to know the processes and reasons why an idea is adopted and others are rejected. Teaching correct and incorrect ideas in parallel is also crucial to present the nuances of each issue, encourage the thorough examination of all alternatives, concretize reflection and prevent indoctrination (Vázquez, Manassero, Acevedo, & Acevedo, 2007). Finally, teaching NoST imposes consistency on the whole curriculum, which will always have to be highly coherent with adequate conceptions of NoST.

Most research on the effectiveness of NoST teaching has been conducted in English-speaking contexts and has involved students in initial training for science teaching (Khishfe, 2008). The study reported in this paper with a group of young science students in a Central American university is part of a broader research project that addresses this topic. It uses a new methodology for empirically studying the effectiveness of teaching the social implications of mining through a short teaching–learning sequence (TLS) about some aspects of the influence of mining on society, and vice versa (the influence of society on mining) and mining technologies (Vázquez-Alonso, Manassero-Mas, & Bennàssar-Roig, 2012).

Research methodology

The research conformed to a pre-test and post-test design with control group. An educational treatment (implementation of the TLS on mining) took place in the experimental group between the two assessments.

Participants

The experimental group consisted of 15 students (9 men and 6 women, average age of 18.7 years) enrolled in an introductory course in chemical sciences in a Central American university. The control group included 25 students (10 men and 15 women, average age of 18.4 years) from the same program as those of the experimental group. These two natural groups of students were similar to each other and selected at random among the available groups. The chemistry teacher of both groups was responsible for implementing the TLS on mining in the experimental group and applying the assessment tool to both groups.

Instruments

The tools used included the TLS on mining, as an educational intervention, and the assessment tool.

The TLS entitled 'Metal extraction: Necessity or nonsense?' is about the social impact of mining activities, on the one hand, and the influence of society on mining, on the other. The innovative character of this TLS lies in the fact that it takes into account the multiple perspectives – scientific, technological, economic, social and environmental – of mining, in such a way that it offers a comprehensive view of the bidirectional impact of S&T on society, and vice versa (see a summary outline of the TLS in the appendix). The TLS structure was based on the general design and common grounds prepared by the broader project's research team from specialized literature (Vázquez et al., 2012).

Both quantitative and qualitative assessment tools were used to measure the effectiveness of the TLS in improving the students' understanding of NoST. Owing to their novelty, utility for future research and space limitations, this study only presents the results from a

Table 1. Items used as pre-test and post-test to assess the effectiveness of the TLS.

VOSTS key	Issue	Item stem
P20211	Companies and research	Scientific research would be better off in our country if it were more closely controlled by corporations (for example, high-technology, communications, pharmaceutical, forestry, mining or manufacturing companies).
P20511	Influence of education Students	The success of science and technology in our country depends on us having good scientists, engineers and technicians. Therefore, our country should require students to study more science in school.
P20521	Influence of education Learning	The success of science and technology in our country depends on how much support the public gives to scientists, engineers and technicians. This support depends on high-school students – the future public – learning how science and technology are used in our country.
P40142	Social responsibility Informing the authorities	When engineers come upon what might be a dangerous idea or product in their work, they actually do inform the public authorities, no matter if it means losing their job or being demoted.
P40311	Balance between pros and cons	We always have to make trade-offs (compromises) between the positive and negative effects of science and technology.
P40511	Promotion of welfare Wealth	The more science and technology develop in our country, the wealthier it will become.
P40521	Promotion of welfare Work	High-technology industries will provide most of the new jobs in the next twenty years.
P40531	Promotion of welfare Quality of life	More technology will improve our country's standard of living.
P80211	Control of technology	Technological development can be controlled by citizens.

standardized paper-and-pencil questionnaire of nine items (Table 1), which were taken from the 'Views on Science-Technology-Society' (VOSTS) questionnaire. VOSTS is a pool of over one hundred empirically developed multiple-choice items covering a large number of issues about NoST (Aikenhead & Ryan, 1992).

For several years, psychometric improvements have been introduced into VOSTS, resulting in a new model of more informative and thorough multiple-response questionnaire, as well as a set of standardized indices that enable the use of inferential statistics (Bennàssar et al., 2010; Vázquez, Manassero, & Acevedo, 2006).

The nine items selected from VOSTS to assess the effectiveness of the TLS on mining were those that were related to the role of mining regarding S&T and its impact on society and that enabled the assessment of the contents and objectives of the TLS. All the items shared a common format: a scenario stem setting out the issue, followed by several statements proposing different rational positions on the issue. Despite sharing the same format,

Table 2. Whole text of the assessment item 20211.

20211 Scientific research would be better off in our country if it were more closely controlled by corporations (for example, high-technology, communications, pharmaceutical, forestry, mining or manufacturing companies).
Corporations should mainly control science:
A. because closer control by corporations would make science more useful and cause discoveries to be made more quickly, through faster communication, better funding and more competition.
B. in order to improve the cooperation between science and technology, and thus solve problems together.
C. but the public or government agencies institutions should have a say in what science tries to achieve.
Corporations should NOT control science:
D. because if corporations did, scientific discoveries would be restricted discoveries that benefit the corporation (for example, the discoveries leading to a profit).
E. because if corporations did, corporations would obstruct scientists from investigating important problems which the companies wanted kept quiet (for example, pollution by the corporation).
F. because the important and transcendent scientific discoveries, which benefit citizens, require doing science without limitations.
G. Science cannot be controlled by corporations. No one, not even the scientist, can control what science will discover.

the items differed in scenario, stem's wording, number of different positions presented and value of the rationales expressed in the statements (see example in Table 2).

The Multiple Response Model (MRM) requires that the person answering the questionnaire evaluate, on a nine-point scale, his or her level of agreement or disagreement with each and every statement for each item setting out the issue. Compared to the Unique Response Model where a single statement is selected in each item among the list of proposed statements, the MRM represents an improvement. By providing the respondent's opinion on every statement, indeed, the MRM maximizes the information available to assess his or her understanding of each issue (Vázquez & Manassero, 1999).

A metric then transforms the raw scores into standardized and normalized indices in the interval [−1, +1]. Indices were calculated from the raw values, taking into account a three-category scaling (Appropriate _A_, Plausible _P_ or Naïve/Ingenuous_N), with each statement previously assigned to one of the three categories by a panel of expert judges according to the current knowledge of specialists in NoST. These indices are the basic quantitative indicators of the respondents' beliefs and link the raw scores, assigned by the respondents, to the categories assigned by the judges to the same VOSTS statements. The more positive and the closer to the maximum value (+1) an index is, the more appropriate and informed a belief is considered; the more negative and the closer to the minimum value (−1) an index is, the more naïve or uninformed the belief is considered. Although the methodology used was quantitative, it enabled and provided the bases for interesting qualitative analyses (Vázquez et al., 2006).

Procedure

The teacher prepared and applied the educational intervention tool (that is, the TLS on mining) to the experimental group in order to teach the features of NoST (see the appendix).

The quasi-experimental pre-test/post-test implementation process was made up of three stages: (i) an initial (pre-test) assessment, conducted by applying the nine-item VOSTS assessment tool; (ii) the application in class of the treatment, that is, the TLS on mining, a month and a half after the initial assessment; and (iii) a final (post-test) assessment using the same tool as for the initial assessment, a month and a half after the treatment. Students were blinded to the experiment and the teacher did not use in class the assessment items. Also, the elapsed time between the pre-test and the post-test was long enough to allow significant effects of teaching to be measured. The control group did not receive the experimental treatment.

The treatment's effectiveness was assessed by comparing, on the one hand, the pre-test assessment's results with those of the post-test assessment, that is, by comparing the scores obtained by the students before and after the TLS was implemented in the experimental group, and, on the other hand, by comparing the experimental group to the control group. Comparisons were made by standardized assessment tools and procedures (Benàssar et al., 2010). The criterion for identifying the most relevant differences between scores and groups was based on the effect size statistic, which enables the magnitude of the difference between two scores to be measured in standard deviation units. By applying this criterion, an effect size was considered relevant if it was larger than 0.30 ($d > .30$), although below this value, many differences could still be statistically significant ($p < .01$).

Results

The objective of this study is to assess improvement in the students' NoST understanding by comparing the variation in scoring of the variables obtained by the assessment tools applied. Analyses focused on statement indices and the weighted average indices for each of the nine VOSTS items, which provide a more global assessment of NoST understanding.

Experimental group: descriptive statistics

Figure 1 shows the average scores of the experimental group – the students who received education on NoST – for the nine VOSTS items, before and after the TLS was implemented. The figure enables us to see the changes that have occurred and answer the question as to whether the treatment had a significant effect on NoST understanding ('effect size' line).

The results about the students' ideas pre-existing to the implementation of the TLS (represented in Figure 1 by the 'pre-test' line) show two items with negative scores (P40142, Social responsibility: Informing authorities, and P40311, Balance between pros and cons), which indicate uninformed pre-existing ideas, and two items with clearly positive average scores (P20521, Influence of education: Learning and P40511, Promotion of welfare: Wealth). The remaining five items show average scores that are positive, but not as much as the two just mentioned. Thus, the students' general pre-test profile is characterized by two items that could be seen as strengths, two as weaknesses and the other items with intermediate scores.

The students' final ideas, after the TLS was implemented (post-test), show a different profile. The two items with initially negative scores are P40142 (Social responsibility: Informing authorities) and P40311 (Balance between pros and cons), which show

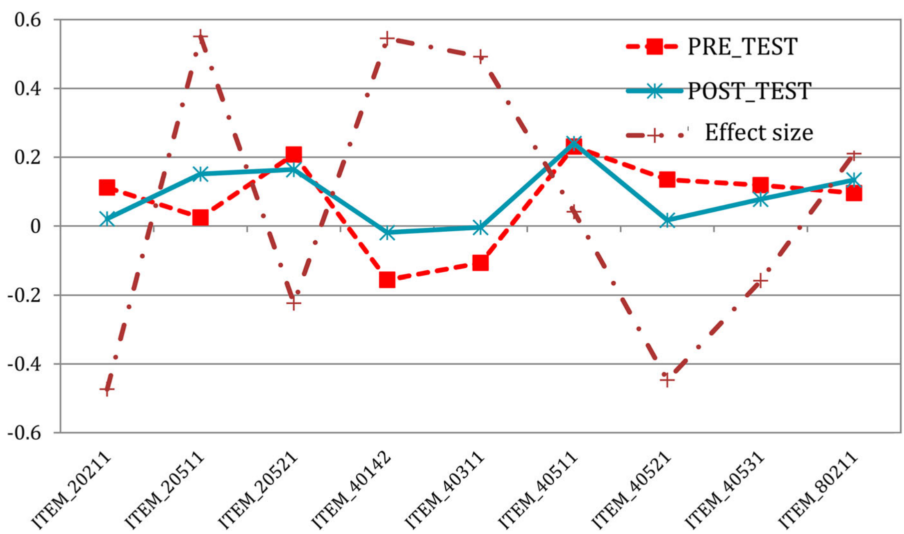

Figure 1. Average item indices for the experimental group in the pre-test and post-test, and effect size of index differences between the two tests.

improved but near-zero scores. The two items with clearly positive average scores (P20521, Influence of education: Learning, and P40511, Promotion of welfare: Wealth) remain without any significant change. Of the five remaining items, three (P20511, P40531 and P80211) maintain positive scores and two (P20211 and P40521) show positive but near-zero scores. The students' general post-test profile shows four items that can be seen as strengths and four other items with near-zero scores.

Performing similar analyses for each of the 59 statements contained in the nine assessed items (Figure 2), the results show 14 statements (24%) with very high average indices (>0.30), that is, strengths that indicate a meaningful understanding of the ideas expressed in these statements. Most of these statements (8) are categorized as appropriate. At the other end of the spectrum, 12 statements (20%) have very negative indices (<0.30), indicating major weaknesses in the students' understanding.

The post-test profile shows 17 statements (29%) with average indices representing strengths in understanding, with most of these statements (nine) categorized as appropriate. At the other end of the spectrum, nine statements (15%) have very negative indices (<0.30), indicating major weaknesses in the students' understanding.

Analysis of improvement: pre–post-test comparisons in the experimental group

The analysis of improvement in NoST understanding was performed by using two quantitative comparison criteria based on the student groups' average indices. First, the differences in the experimental group's understanding before and after the TLS was implemented are compared, and second, the differences between the control and the experimental group are analyzed.

For the experimental group, the average index differences before and after the TLS was implemented are also analyzed through various indicators related to the effect size calculated from the item and statement indices. The sum of the average effect sizes along the nine items indices – a positive effect size indicating a gain in understanding and a negative one, a weakened understanding – showed a positive value (+0.135), which indicates that the experimental group's overall understanding of NoST has improved after the TLS.

The pre–post-test differences in the understanding of the nine items are represented in Figure 1 by the 'effect size' line, with positive values indicating improvements. A closer look at Figure 1 reveals that the most important improvements ($d > .3$) occurred in three items (P20511, P40142 and P40311). Two items (P20211 and P40521) showed a decline, and for the other four, differences were considered irrelevant ($d < .3$). The differences were not statistically significant in any case (Mann–Whitney U test).

The same pre–post-test comparisons of the average indices for the experimental group was performed for each of the 59 statements and are shown in Figure 2 by the 'effect size' dotted bars.

The sum of the average effect sizes along all the statement indices showed a zero value, which indicates that overall, the differences in statement understanding before and after the TLS compensate each other, that is, the average magnitude of improvements matches that of declines.

However, when the percentage of statements representing strengths in understanding was compared before (25%) and after (21%) the TLS was implemented, and similarly for

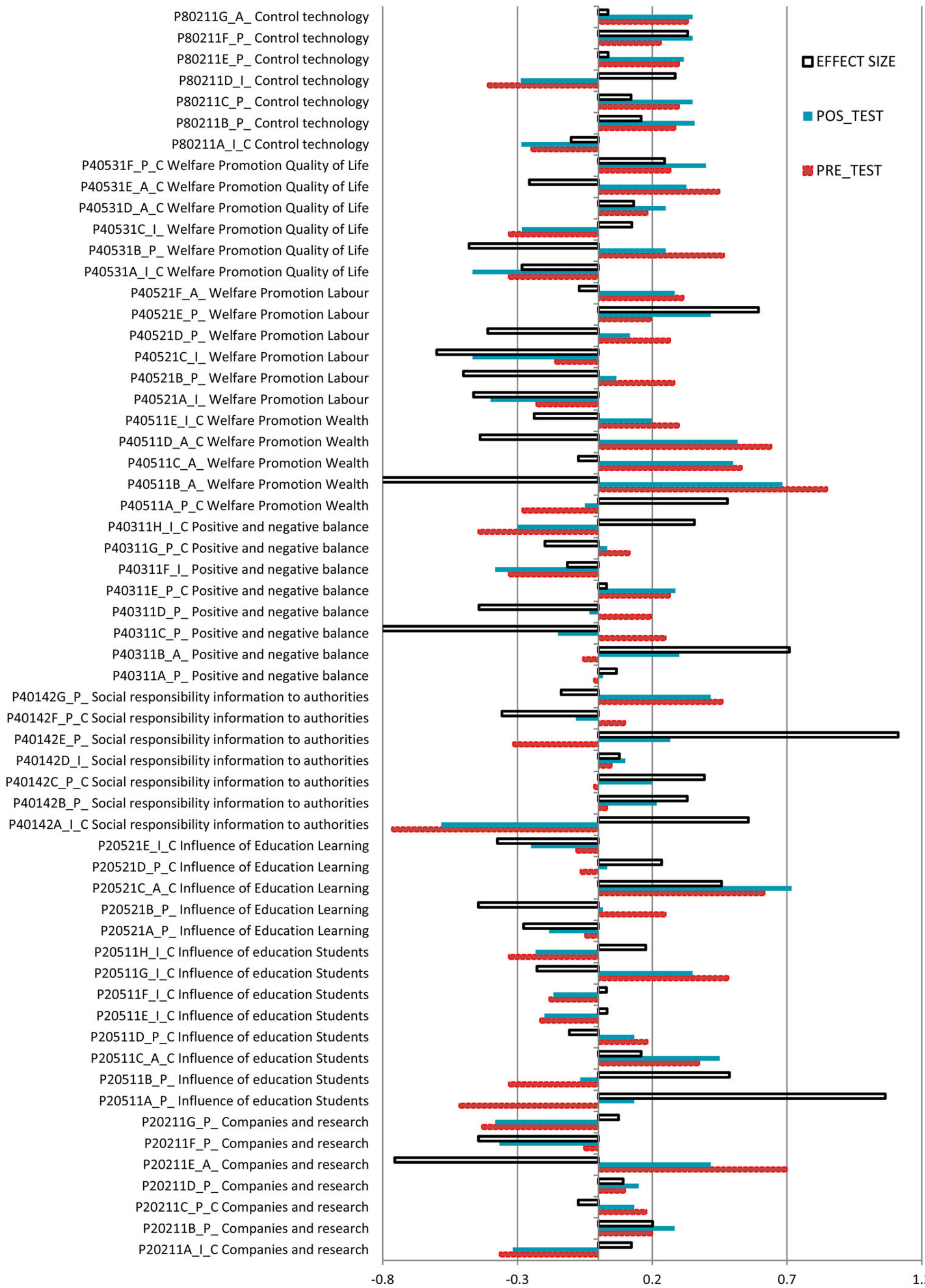

Figure 2. Average statement indices for the experimental group in the pre-test and post-test, and effect size of index differences between the two tests.

Table 3. Average pre-test and post-test statement indices associated with significant effect sizes, used to assess the effectiveness of a TLS on mining.

	Pre-test			Post-test			
Statement	Sample size	Average	Standard deviation	Sample size	Average	Standard deviation	Effect size of differences
P20511A_P_ Influence of education: students	15	−0.517	.671	15	0.133	.550	+1.065
P40142E_P_ Social responsibility: informing the authorities	15	−0.317	.555	15	0.267	.495	+1.112
P40311C_P_ Balance between pros and cons	15	0.25	.463	15	−0.15	.533	−0.803
P40511B_A_ Promotion of welfare: wealth	15	0.85	.158	15	0.683	.240	−0.838

the percentage of statements representing weaknesses (18% before vs. 13% after), a tendency to improvement could be observed.

Of the 59 pre–post-test comparisons for each of the statements, only four produced statistically significant differences (Mann–Whitney *U* test). Two statements showed positive differences (that is, an improvement in understanding), while the other two displayed negative differences (a decline in understanding) (Table 3).

In short, the pre–post-test effect sizes of the differences confirmed that the TLS was a determinant of change in the experimental group's understanding of NoST, as they showed a tendency to improvement, though of modest magnitude: an improvement in three items, a modest decline in two items, and no relevant change in four items. The improvement trend was much clearer when measured by overall changes in the 59 statements (Figure 2), as the number of statements representing strengths – whose understanding was very meaningful – increased (+4%), while the number of statements representing weaknesses – exhibiting a very poor understanding – decreased (−3%).

Analysis of improvement: comparisons between the experimental and the control group

Comparing the differences in understanding between the control and the experimental group is another effective way of assessing the impact of the TLS on improving NoST understanding. Although both groups were equivalent – though non-explicit, potential intervening variables were shared in both groups – their starting points of understanding (pre-test indices) were not equal, if only because of mere sampling variation.

Figure 3 displays this initial variability within the two groups, which did not agree on most items. Two criteria could be used to determine which group improved the most: an absolute-difference criterion, which simply considered which group had the best understanding, and a relative-difference criterion, which complemented the absolute criterion by controlling for the different starting points at NoST understanding. For each item, the relative criterion for best improvement was the highest (most positive) slope of the straight-line pair for both groups (Figure 3). Regardless of their starting points, when the two lines were nearly parallel, it meant that there was no difference between groups; otherwise, the line with the highest slope determined the group with the greatest improvement. According to this criterion, five items (20511, 40142, 40311, 40511 and 80211)

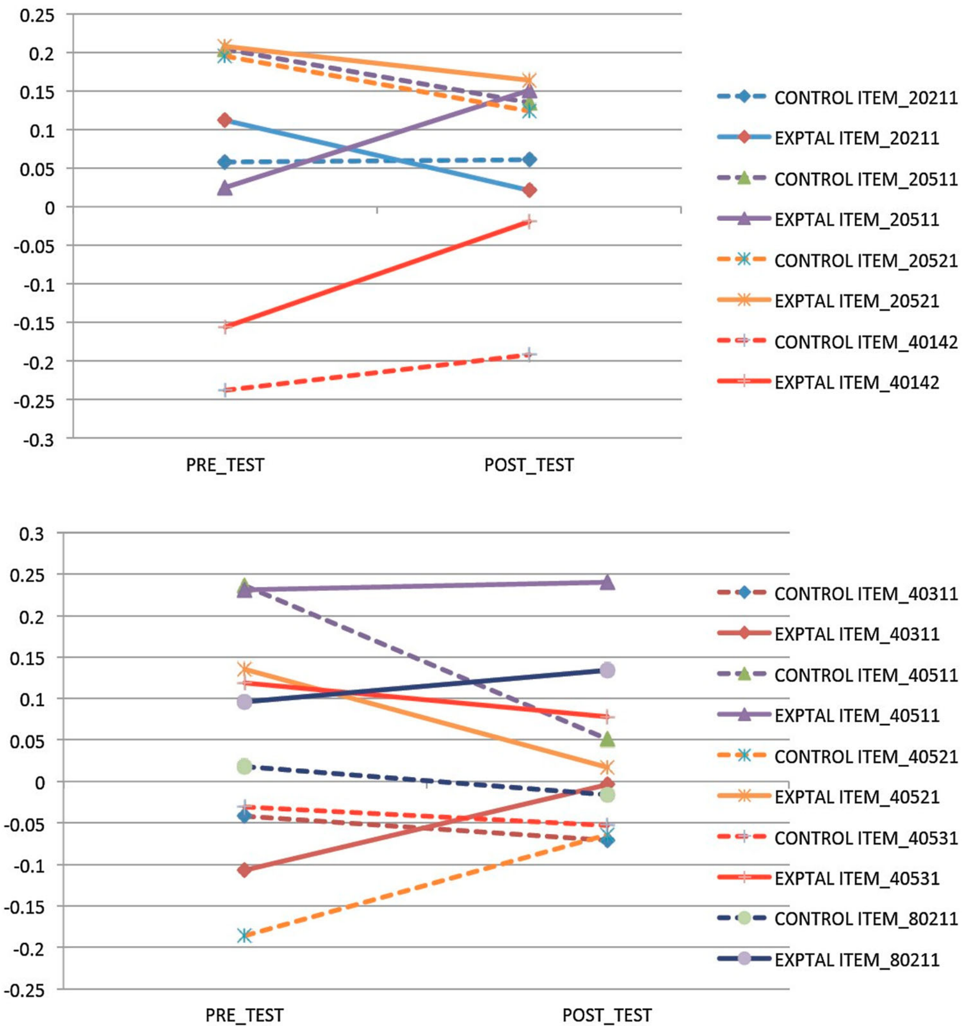

Figure 3. Average indices of the nine items for the experimental (solid lines) and control (dotted lines) groups before ('pre-test') and after ('post-test') the implementation of the TLS on mining.

showed greater improvement in the experimental group, two items (20521 and 40531) showed similar improvement in both groups and the two remaining items (20211 and 40521) showed greater improvement in the control group.

Figure 4 summarizes these results by showing, for each item, the effect size of the average difference between the pre-test and the post-test in the experimental and the control group, hence enabling the magnitude of the change to be compared between the two groups for each item. The experimental group achieved better understanding than the control group in most items, so that the balance of this comparison favors the experimental group.

The final analysis performed can be expressed as follows: From the results at the end of the learning process (post-test average indices), has the overall NoST understanding of the experimental group improved with respect to the control group?

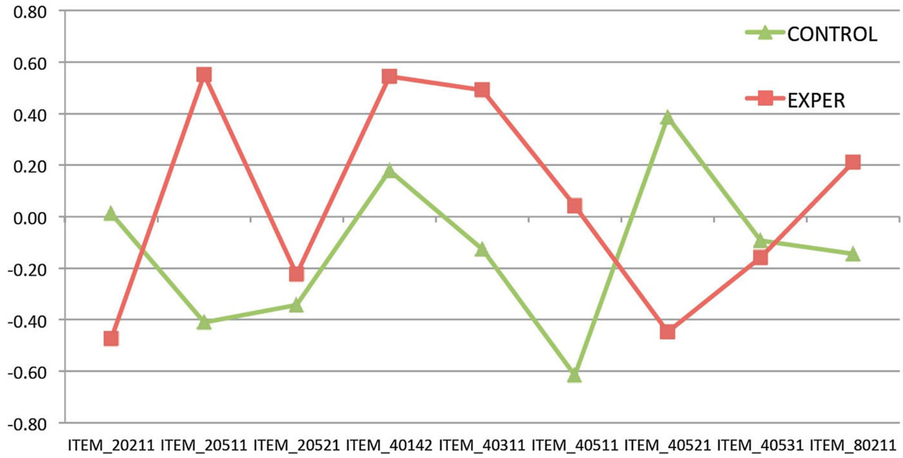

Figure 4. Effect size of the pre–post-test differences in the experimental and the control group for all items.

Figure 5 compares, for all nine items, the final (post-test) average indices between the two groups. It displays the effect size of index differences of the experimental group with respect to the control group, with positive effect sizes indicating higher indices in the experimental group. It can be seen that the experimental group obtained better final average indices than the control group on all items except one (20211), and the effect size of these differences with respect to the control group was relevant and significant ($d > .30$) for six items. It thus seems that by the end of the TLS, the experimental group had achieved a much better understanding than the control group for most NoST issues.

When the previous comparison analysis between the experimental and the control group was replicated along the average indices of the 59 statements, the overall result was also mostly favorable to the experimental group (Figure 6).

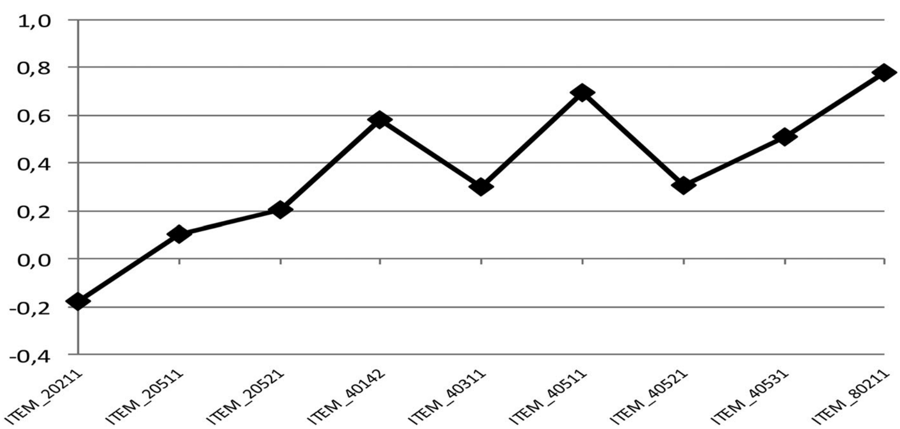

Figure 5. Effect size of the differences in post-test average item indices between the experimental and the control group (a positive effect size indicates higher index for the experimental group).

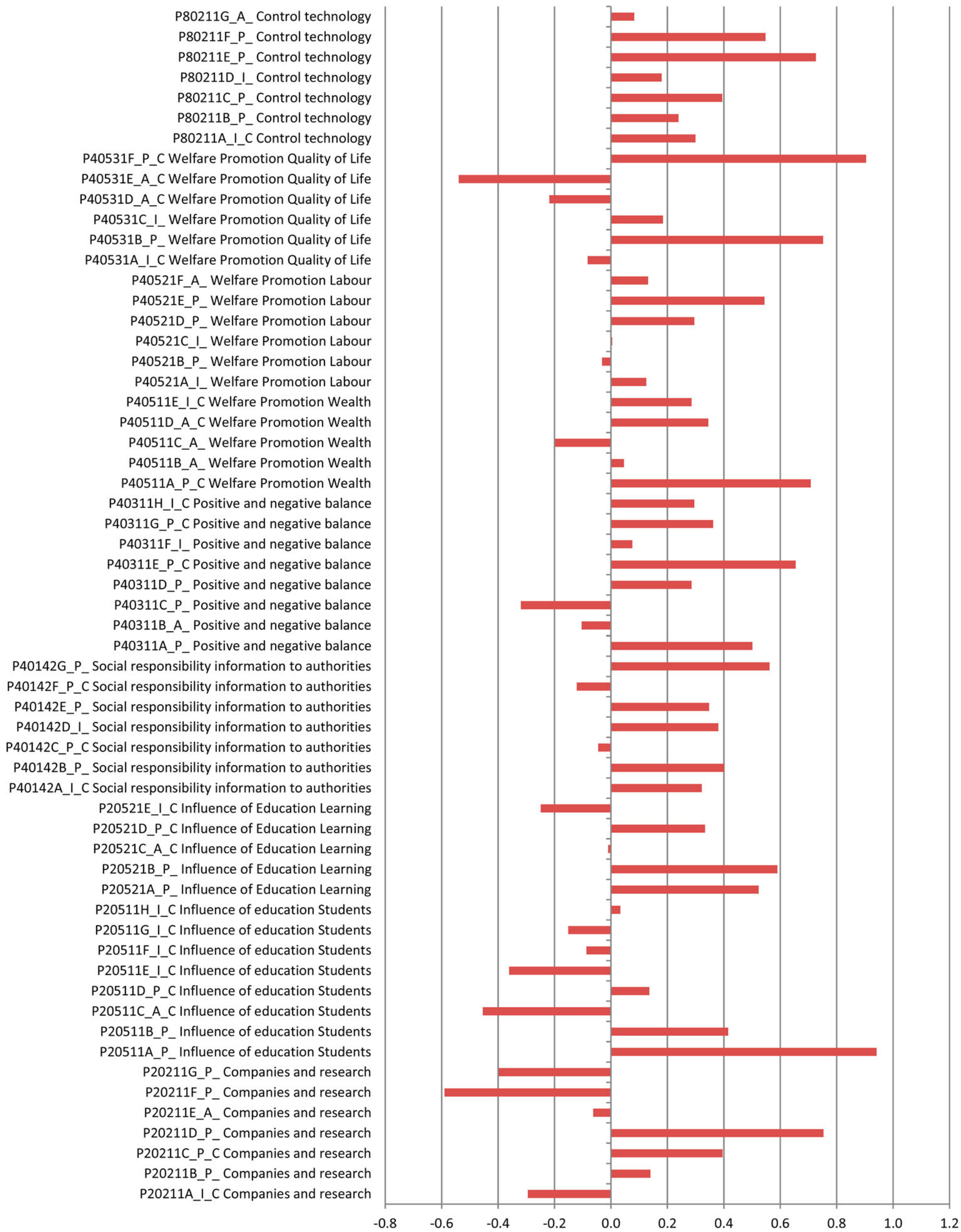

Figure 6. Effect size of the differences in post-test average statement indices between the experimental and the control group (a positive effect size indicates higher index for the experimental group).

For the great majority of statements (40), the experimental group obtained better average scores than the control group. Moreover, these better final indices of the experimental group with respect to the control group were relevant and significant ($d > .30$) in 26 statements, while the reverse (significantly better indices of the control group, $d < .30$) occurred in only seven statements. These results regarding statement indices show once again that at the end of the TLS, the experimental group achieved a significantly better NoST understanding than the control group.

Overall, comparing the experimental group to the control group served to control for the influence of the multiple intervening factors that were unknown or not evaluated. Comparing the pre–post-test improvements achieved by the experimental and the control group showed that the first group had significantly better improvement than the second in five items, although the reverse also occurred in two items. Comparing the two groups' final item indices was more convincing: the experimental group obtained better average indices than the control group for all items except one, with relevant and significant differences ($d > .30$) in six items.

Compared in a similar way, the final average indices of the 59 items were also clearly favorable to the experimental group, which obtained better average scores than the control group for the great majority of statements (40). Moreover, these differences with respect to the control group were relevant and significant ($d > .30$) in 26 statements – almost half of them, while the reverse (control group significantly better than the experimental group) occurred in only seven statements.

This quantitative assessment translates into formative assessment through the qualitative identification of the specific strengths and weaknesses of the explicit and reflective teaching on mining. For brevity, only the weaknesses are listed below, which correspond to the following statements (sorted by decreasing difference to the experimental group):

- P20211F_P_ Companies and research ('… Corporations should not control science: because the important and transcendent scientific discoveries, which benefit citizens, require practicing science without any constraints.')
- P40531E_A_C Promotion of welfare: Quality of life (Yes and no. 'More technology would make life easier, healthier and more efficient. BUT more technology would cause more pollution, unemployment and other problems. The standard of living may improve, but the quality of life may not.')
- P20511C_A_C Influence of education: Students ('… Students should be required to study more science: because it is important to help our country to keep up with other countries.')
- P20211G_P_ Companies and research ('… Corporations should not control science: Science cannot be controlled by corporations. No one, not even the scientist, can control what science will discover.')
- P20511E_I_C Influence of education: Students ('Students should not be required to study more science: because it won't work. Some people don't like science. If you force them to study it, it will be a waste of time and will turn people away from science.')
- P40311C_P_ Balance between pros and cons ('… because things that benefit some people will be negative for someone else. This depends on a person's viewpoint.')

Despite these weaknesses, the above results show that the experimental group, at the end of the TLS, achieved a significantly better understanding of NoST issues than the control group. The result about the final achievement along the statements is even more clearly favorable to the experimental group.

Conclusions

This study examines the effectiveness of a TLS to improve understanding of NoST by using an example of socio-scientific issue: the mutual influence of a technology (mining) and society. The contributions of the study focus on the transfer to the classroom of all the materials (TLS and assessment tool), thanks to the standardized experimental and instructional design (pre–post-test design with control group), which enables not only non-specialist teachers to face the challenge of teaching NoST, but also to compare the effects of different didactic sequences or different studies. Both aims are today difficult to achieve because of the incommensurable methods used by different researchers (Abd-El-Khalick & Lederman, 2000; Rudge & Howe, 2009).

The improvement in NoST understanding is evaluated by comparing the average indices of 59 statements (grouped in nine assessment items) before and after the implementation of the TLS in the experimental group, and between the experimental and the control group. First, the magnitude of the experimental group's improvement through the TLS was clear, modest for items but greater for individual statements, and allowed strengths and weaknesses in understanding to be identified. The identified strengths are related to the issues of social responsibility: informing the authorities (40142) and balance between pros and cons (40311), while the weaknesses occurred in the issues of companies and research (20211) and promotion of welfare and work (40521).

Second, the experimental group exhibited greater improvement than the control group in eight of the nine issues, that of companies and research (20211) being the only one that did not improve. It should be noted, however, that this improvement was not significant for the issue of the influence of education to recruit science students (20511). It should be stressed that the interpretation of the indices' small numerical values (lying in the [−1, +1] interval) and of their differences should not be intuitively made based on the small absolute number values. Rather, their correct interpretation is offered by the effect size, a statistic measuring differences at a normal scale (in standard deviation units), making these difference values independent of the original-scale numerical values.

In short, the assessment of TLS effectiveness made by comparing the experimental to the control group showed that the former achieved a significantly better NoST understanding. This significant, though modest, improvement replicates the findings of other studies, where improvements, though real, were not large or did not affect all the aspects taught (Acevedo, 2009; Deng et al., 2011; Rudge and Howe, 2009). To better appreciate these modest gains in NoST understanding, it must be kept in mind that the students were blinded to the educational experiment (the implementation of a TLS on mining). The educational intervention was simple, straightforward and of short duration, and its main innovation consisted in explicitly addressing, and having students reflect on, the mutual impact of mining and society (see the TLS activities in the appendix). Therefore, the students' improvements in NoST understanding are genuine, internalized and meaningful: they are not attributable to the external stimuli that usually motivate learning (passing exams, obtaining good grades, etc.), but to intrinsic motivation, to having examined, debated and used argumentation about a socio-technical issue, the role of the mining industry.

A detailed analysis of the assessment (average indices of the statements making up each item before and after the TLS) resulted in a genuinely formative assessment to improve

NoST understanding and teaching: a profile for each student and group showing strengths, weaknesses, opportunities and challenges, hence allowing for formative self-reflection about NoST learning. Such a detailed analysis helps us to determine which specific statements were responsible for decreased NoST understanding and to design a recovery strategy to clarify those issues. For example, the decreased understanding of statement 40511B ('… because more science and technology would make our country less dependent on other countries. We could produce things ourselves.') from a very high average score (appropriate) to a lowest but still high score (Table 3) could be attributed to the critical thinking developed through the TLS, which could modulate the students' initially high confidence in their beliefs towards a more critical view of the role of technology in promoting welfare. Further examples are item 40521, where the decrease in understanding is due to only the first four statements, and item 20211, where it is caused by only two statements. Such a formative, personalized and empirically based assessment thus allows producing genuine education, reinforcing strengths and improving weaknesses in understanding.

Finally, this research approaches NoST teaching from a different perspective than most previous studies and thus differs in various aspects. First, instead of addressing various aspects of NoST at the same time or a broader NoST issue, it focuses on a specific and simple issue (the socio-technical role of mining), with a view to design assessment tools that are specific to the chosen aspect. By eliminating excessive and superfluous details, this approach makes it easier to relate NoST with specific scientific concepts (in this case, the chemistry of metals), learning objectives and students' ideas in the classroom (Ryder & Leach, 2009; Taber, 2008). The specificity of the approach provides more context and reduces the innate NoST complexity, which has positive effects on students' motivation and teachers' confidence to give knowledge of NoST without special preparation (Niaz, 2009).

Second, unlike the studies presented in the Introduction, whose assessments are mostly based on open-ended, qualitative students' productions that lead to few idiosyncratic and very general conclusions, the standardized quantitative approach used here naturally produces empirical data that lead to clear conclusions and also enable qualitative analyses (strengths and weaknesses, significant changes, personal progress profiles, transition or stagnation, etc.). These standardized data could be complemented by open-ended students' productions (Abd-El-Khalick & Lederman, 2000) as part of the TLS activities, such as debates, conclusions and reports (see description in the appendix), which are not presented here due to space limitations.

Third, a no less important contribution of this study is the standardized assessment tool adapted to each specific issue, as the lack of assessment criteria and instruments often prevents teachers from teaching NoST. The availability of an instrument that is standardized (does not require specialized knowledge for use), is flexible (customizable), produces individualized profiles and can be adapted to different needs, may give teachers confidence and may overcome barriers to teaching NoTS. At the same time, this standardized tool allows comparisons to be made between different NoST educational treatments as well as between the results of different studies related to NoST teaching, which could not be compared until now due to the use of open-ended tests. This tool makes it possible to overcome the lack of reproducibility and comparability among studies on NoST teaching, which has slowed down progress in the area (Khishfe, 2008). Further research on other teaching cases is, of course, required to confirm the effectiveness of the line of research presented here, with a view to promote better NoST teaching by developing teacher training in pedagogical content knowledge and in the area of formative assessment (Abd-El-Khalick & Akerson, 2009; Hanuscin, Lee, & Akerson, 2011).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Research Project EDU2010-16553 funded by the National RDI Plan of Spain's Ministry of Science and Innovation.

Notes on Contributors

Ángel Vázquez-Alonso is a senior researcher, teacher in the master degree for training high-school science teachers and leader of the research project EDU2010-16553.

Abdiel Aponte is an associate Chemistry teacher and researcher of the research project EDU2010-16553.

María-Antonia Manassero-Mas is a full Professor of Social Psychology and leader in several peer-evaluated research projects and main researcher of the research project EDU2010-16553.

Marisa Montesano was chief of the National Secretary for Science, Technology and Innovation (SENACYT) of the Panamanian Government, and coordinator for Panamá of the research project EDU2010-16553.

References

Abd-El-Khalick, F., & Akerson, V. (2009). The influence of metacognitive training on preservice elementary teachers' conceptions of nature of science. *International Journal of Science Education*, *31*, 2161–2184.

Abd-El-Khalick, F., & Lederman, N. G. (2000). Improving science teachers' conceptions of nature of science: A critical review of the literature. *International Journal of Science Education*, *22*(7), 665–701.

Acevedo, J. A. (2009). Enfoques explícitos versus implícitos en la enseñanza de la naturaleza de la ciencia. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, *6*(3), 355–386.

Acevedo, J. A., Vázquez, A., Manassero, M. A., & Acevedo, P. (2007). Consensos sobre la naturaleza de la ciencia: aspectos epistemológicos. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, *4*(2), 202–225.

Aikenhead, G. S., & Ryan, A. G. (1992). The development of a new instrument: 'Views on Science-Technology-Society' (VOSTS). *Science Education*, *76*(5), 477–491.

Bennàssar, A., Vázquez, A., Manassero, M. A., & García-Carmona, A. (Eds.). (2010). *Ciencia, tecnología y sociedad en Iberoamérica: una evaluación de la comprensión de la naturaleza de ciencia y tecnología.* Madrid: OEI. Retrieved from http://www.oei.es/salactsi/DOCUMENTO5vf.pdf

Clough, M. P. (2007). Teaching the nature of science to secondary and post-secondary students: Questions rather than tenets. *The Pantaneto Forum*, *25*, Retrieved from http://www.pantaneto.co.uk/issue25/front25.htm

Deng, F., Chen, D.-T., Tsai, C.-C., & Chai, C. S. (2011). Students' views of the nature of science: A critical review of research. *Science Education*, *95*, 961–999.

García-Carmona, A., Vázquez, A., & Manassero, M. A. (2012). Comprensión de los estudiantes sobre naturaleza de la ciencia: análisis del estado actual de la cuestión y perspectivas. *Enseñanza de las Ciencias*, *30*(1), 23–34.

Hanuscin, D. L., Lee, M. H., & Akerson, V. L. (2011). Elementary teachers' pedagogical content knowledge for teaching the nature of science. *Science Education*, *95*(1), 145–167.

Hodson, D. (2008). *Towards scientific literacy: A teachers' guide to the history, philosophy and sociology of science*. Rotterdam: Sense.

Hodson, D. (2009). *Teaching and learning about science: Language, theories, methods, history, traditions and value*. Rotterdam: Sense.

Khishfe, R. (2008). The development of seventh graders' views of nature of science. *Journal of Research in Science Teaching*, *45*(4), 470–496.

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: Lawrence Erlbaum Associates.

Matthews, M. R. (1994). *Science teaching: The role of history and philosophy of science*. London: Routledge.

Matthews, M. R. (1998). In defense of modest goals when teaching about the nature of science. *Journal of Research in Science Teaching*, *35*, 161–174.

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. Concepts and methodologies* (pp. 3–26). Dordrecht: Springer.

McComas, W. F., & Olson, J. K. (1998). The nature of science in international science education standards documents. In W. F. McComas (Ed.), *The nature of science in science education: Rationales and strategies* (pp. 41–52). Dordrecht: Kluwer Academic.

Millar, R. (2006). Twenty first century science: Insights from the design and implementation of a scientific literacy approach in school science. *International Journal of Science Education*, *28* (13), 1499–1521.

Niaz, M. (2009). Progressive transitions in chemistry teachers' understanding of nature of science based on historical controversies. *Science & Education*, *18*, 43–65.

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.

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.

Ryder, J., & Leach, J. (2009). Teaching about the epistemology of science in upper secondary schools: An analysis of teachers' classroom talk. *Science & Education*, *18*, 43–65.

Taber, K. (2008). Towards a curricular model of the nature of science. *Science & Education*, *17*, 179–218.

Tala, S. (2009). Unified view of science and technology for education: Technoscience and technoscience education. *Science & Education*, *18*, 275–298.

Vázquez, A., & Manassero, M. A. (1999). Response and scoring models for the 'Views on Science-Technology-Society' instrument. *International Journal of Science Education*, *21*(3), 231–247.

Vázquez, A., & Manassero, M. A. (2008). Concepciones de profesores en formación inicial sobre naturaleza de la ciencia y la tecnología. *Tecnología y Cultura*, *10*(13), 18–28.

Vázquez, A., Manassero, M. A., & Acevedo, J. A. (2006). An analysis of complex multiple-choice science–technology–society items: Methodological development and preliminary results. *Science Education*, *90*(4), 681–706.

Vázquez, A., Manassero, M. A., Acevedo, J. A., & Acevedo, P. (2007). Consensos sobre la naturaleza de la ciencia: la ciencia y la tecnología en la sociedad. *Educación Química*, *18*(1), 38–55.

Vázquez-Alonso, Á, & Manassero-Mas, M. A. (2012). La selección de contenidos para enseñar naturaleza de la ciencia y tecnología (parte 1): Una revisión de las aportaciones de la investigación didáctica. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, *9*(1), 2–33.

Vázquez-Alonso, Á, Manassero-Mas, M. A., & Bennàssar-Roig, A. (2012). *Proyecto EANCYT: Enseñar, aprender y evaluar sobre naturaleza de la ciencia y tecnología* [EANCYT project:

Teaching, learning and assessment of nature of science and technology]. Paper presented at the First International Symposium on Science Teaching (I ISIEC 2012).
Zeidler, D. L., Sadler, T. D., Simmons, M. L., & Howes, E. V. (2005). Beyond STS: A research-based framework for socioscientific issues education. *Science Education*, *89*(3), 357–377.

Appendix

LEARNING SEQUENCE

'Metal extraction: Necessity or nonsense?' (6 lessons to students enrolled in the first year of the Chemistry degree, Content Module on the 'Solid State')

The approach meets the teacher's concern that the student must know the features of mining in all its possible aspects (scientific, technological, economic, social and environmental), in order for him or her to get a global view of the issue.

OBJECTIVES

1. Recognize, value and understand the features and implications specific to mining, throughout history and nowadays in our country.
2. Develop an analytical and critical view of the implications of mining in the case of Panama.
3. Reflect on the relationships of natural sciences with technology and society.
4. Know and evaluate the social and cultural implications that developing mining can have in Panama.

SCIENTIFIC COMPETENCIES

Support opinions, identify problems, interpret and explain processes, determine interrelationships and evaluate preconceptions

ACTIVITIES (Students/Teachers)	Methodology/ Organization	Materials/Resources
ENGAGE Introduction-motivation Where are metals extracted from? Give examples Do you know what are the processes for obtaining a pure metal? / Tell what you know about our mineral wealth What interests have driven renewed interest in mining in Panama?	Brainstorming The whole group	Videos on pros and cons Newspaper's news
EXTRACT Prior knowledge The students will answer the following questions: Have minerals already been exploited in Panama? What extraction methods do you think were used? Does mineral extraction pose a risk to the environment? What happens when minerals are not extracted from the earth's crust?	The whole class Workshops	Explanations provided and predictions made Questionnaire
Development Activities		
EXPLAIN Contents Concepts of metallurgy: Separation techniques. Chemical reactions. Chemical properties of metals. Uses and applications. Group search for information on metallurgy and the properties and applications of the metals obtained from these processes. (The students will prepare group reports using the data obtained from searching information). *The teacher assigns reading on past mining in Panama and its consequences so far. Picking up and review of reports.*	Teams of three students Collaborative work	Printed material ICTs Multimedia equipment

(*Continued*)

Appendix Continued.

ACTIVITIES (Students/Teachers)	Methodology/ Organization	Materials/Resources
EXPLAIN EXPLORE Procedures *From the reports previously picked up by the teacher, preparation and review of general documentation for the whole group, which will be worked on in class and serve as a basis for the upcoming visit.* Acknowledge and value the importance that mining and the chemical industry have had in the country's historical, social and economic development. The students write all their observations in the laboratory guide. They perform an experiment of extracting copper from a commercial malachite sample.	Debate and argumentation Small groups (four participants)	
EXPLAIN EXPLORE Attitudes Observation of the social and cultural impacts on the areas surrounding mining sites. Comparison of the knowledge acquired during the visit to the mine with that previously obtained from the group reports.	Group	Recording of attitudes: Observation sheet
EXPLORE Strengthening knowledge *The teacher prepares a questionnaire of the relevant aspects that the student has to notice during the visit.* *The teacher gives the guiding documentation for the visit. Visit to Petaquilla.* *The student makes observations and collects data in situ during the guided visit to the Petaquilla mine.* *The teacher leads the review and strengthening of the concepts previously acquired or mentioned during the visit.* The student debates and analyzes the actual situation and draws conclusions.	Group	Document for the guided visit to the mine
The students will perform a laboratory experience to extract copper from a mineral. *The teacher prepares the framework for writing the report.* The student prepares a report based on the assigned framework.	Practical work	Laboratory guide Commercial malachite sample
Assessment		
Items P20211, P20511, P20521, P40142, P40311, P40511, P40521, P40531 and P80211	Individual	VOSTS items

Note: The teacher's activities are in italics.