**The final publication is available at link.springer.com**

**(https://link.springer.com/article/10.1007%2Fs11251-018-9446-4)**

# Preservice Elementary Science Teachers' Argumentation Competence: Impact of a Training Programme

**Daniel Cebrián-Robles; Antonio-Joaquín Franco-Mariscal; Ángel Blanco-López**

**Antonio-Joaquín Franco-Mariscal anjoa@uma.es**

**Daniel Cebrián-Robles dcebrian@uma.es**

**Ángel Blanco-López ablancol@uma.es**

**Facultad de Ciencias de la Educación, Didáctica de las Ciencias Experimentales, Universidad de Málaga, Campus de Teatinos s/n. 29071, Málaga, Spain**

**Preservice Elementary Science Teachers' Argumentation Competence: Impact of a Training Programme**

**Abstract**: The recent literature has shown the importance of Preservice Elementary Science Teachers (PESTs) having a deep understanding of argumentation, as this factor may affect the nature of the class activities that are taught and what students learn. A lack of understanding of this factor may represent an obstacle in the development of science education programmes in line with the development of scientific competences. This paper presents the results of the design and implementation of a training programme of 6 sessions (12 hours of class participation plus 8 hours of personal homework) on argumentation. The programme was carried out by 57 Spanish PESTs from Malaga, Spain. The training programme incorporates the innovative use of certain strategies to improve competence in argumentation, such as teaching PESTs to identify the elements of arguments in order to design assessment rubrics or by including peer assessment during evaluation with and without rubrics. The results obtained on implementing the training programme were evaluated based on the development of PESTs' argumentation competence using Toulmin's argumentative model. Data collection methods involved two tasks carried out at the beginning and the end of the programme, i.e., pre-test and post-test, respectively. The conclusion of the study is that students made significant progress in their argumentation competence on completing the course. In addition, PESTs who followed the training programme achieved statistically better results at the end than those in the control group (n = 41), who followed a traditional teaching programme. A 6-month transfer task showed a slight improvement for the PESTs of the experimental group in relation to the control group in their ability to transfer argumentation to practice, especially to the extent to which they mentioned argumentation in their practice portfolios.

**Keywords:** Argumentation Competence, Preservice Elementary Science Teachers, Rubrics, Assessment.

## Introduction

Argumentation is now considered to be one of the main scientific practices and it is thus a key element in science teaching (Erduran &, Jiménez-Aleixandre, 2008). McNeil and Knight (2013) consider it necessary to work explicitly on this area with future trainee teachers and teachers in service as the best way to address argumentation in class. In order to design and evaluate argumentative activities, an appropriate model is required to understand argumentation. This is because argumentation in science classes can engage students in heuristic strategies to learn to reason and, at the same time, their arguments act as the outsourcing of their reasoning, thus allowing them to assess and improve their own learning (Duschl, Ellenbogen, & Erduran, 1999; Driver, Newton, & Osborne, 2000; Jiménez-Aleixandre, Bugallo, & Duschl, 2000; Kelly, & Takao, 2002; Osborne, Erduran, & Simon, 2004; Jiménez-Aleixandre, 2005; Justi, 2006).

Despite the interest in improving argumentation skills, different research studies revealed that there are very few chances for students to participate in argumentation activities in science classes (Sampson, & Blanchard, 2012). A number of factors reinforce the above; for instance, the role played by the teacher in bringing argumentation into class. In relation to this aspect, Sampson and Blanchard (2012) stated that:

> Understanding what teachers know about argumentation and their views about the role of argumentation in science education is important because the extent to which a new curriculum, instructional strategy, or technology-enhanced learning environment is actually used depends largely on what teachers know, what they value, and how they decide to use it (p. 1125).

In spite of its importance, there is little knowledge about teachers' understanding of argumentation, their ability to participate in this complex practice or their views on the use of argumentation in the teaching-learning process in science. The available literature suggests that a teacher's knowledge about argumentation can influence the nature of their class activities and what students learn (Andrews, & Mitchell, 2001; Sampson & Blanchard, 2012). The literature also highlights the importance of teachers having a deep understanding of argumentation and arguments not only in elementary classroom community (Kovalainen & Kumpulainen, 2005) but also for middle school students (Belland, Glazewski, & Richardson, 2011; Marttunen, & Laurinen, 2001). According to Starling and Lee (2015) there is a lack of ability to construct arguments or justifications for a point of view within group discussions in

secondary education. Litman et al. (2017) also addressed the particular focus on the argumentation that the secondary teachers should have in order to transform literacy instruction in science education. These authors suggest that intervention designs should foreground the close reading and cross-textual analysis as building blocks of argumentation. In the same way, an extensive body of papers make an effort to present a training programme for high school students improving their argumentative competence (Marttunen, 2002; Marttunen, & Laurinen, 2007; Bulgren, Ellis, & Marquis, 2014; Ragonis, & Shilo, 2014).

This problem cited above is especially prevalent among primary science teachers, who show a low level of scientific knowledge and skills (Andrews, & Mitchell, 2001). On extending this premise to the wider population it is found that there is a lack of an adequate ability to make scientific arguments in everyday situations (Klein, 2004). This situation may represent an obstacle to developing science education programmes in line with the development of scientific competencies, including the argumentation competence (Osborne et al. 2016). As a consequence, in the study reported here it was assumed that argumentation is one of the main objectives that should be taught to both trainee teachers and science students (Driver, Newton, & Osborne, 2000; Duschl, & Osborne, 2002). The authors of this research feel it necessary to influence implicitly an improvement in the argumentation competence in science teachers from their initial training (Archila, 2012). Some training programmes for schools and high schools have been found to improve the argumentation competence (Marttunen, & Laurinen, 2001; Hefter et al., 2014). However, the aim of this study was to assess the impact of a training programme on argumentation competence in Preservice Elementary Science Teachers (hereinafter PESTs).

## Theoretical Framework

### Argumentation in Science and Science Education

Nowadays, argumentation is considered to be a key pillar of science (Duschl, & Osborne, 2002). It is precisely through argumentative processes –conditioned by history– that cultural presumptions and science identities have emerged (Taylor, 1996; Driver et al., 2000). On a separate note, argumentation can also be framed within the concept of scientific practices and these are regarded as essential in scientific work. According to Berland and Reiser (2011), scientific practices are *argumentation* and *modelling*. It is accepted that the building of the scientific mind requires a complex combination of skills, processes, and

attitudes, and among these, argumentation plays an important role (Visser, 2007). In short, taking part in science involves discussing, reasoning, arguing, criticising and justifying ideas and explanations.

The literature contains different definitions for argumentation in science. Jiménez-Aleixandre and Díaz (2003, p.361) consider scientific argumentation to be *"*the ability to relate data and conclusions, to evaluate theoretical propositions in the light of empirical data or data from other sources*"*. However, for Sanmartí (2003) argumentation is*:*

> A social, intellectual and verbal activity that serves to justify or refute an opinion, and consists of making statements taking into account the recipient and the purpose of communication. To argue is then to choose from different options or explanations and to reason the criteria for assessing why the chosen option was the best (p. 123).

In this sense, the process of arguing is to produce a text with the purpose of changing the epistemic value of the arguments held by the recipient, providing significant reasons for such a change in order to persuade the recipient to see that the new ideas are 'justified' by evidence or other sources. Jiménez-Aleixandre and Puig (2010) stated that an argument could not exist without submitting certain scientific knowledge to evaluation and providing evidence or reasons to either confirm or refute this knowledge.

According to Henderson, Osborne MacPherson and Szu (2014), argumentation should not only be seen as an attitude, but rather as a scientific competence, which they call scientific argumentation and define as: *"*a complex form of reasoning demanded by situations that require scientific content knowledge to construct and/or critique proposed links between hypotheses and evidence*"* (p. 2). To these authors, the model for a good argument must have three basic elements: conclusion, justification and evidence – as also required by Toulmin's model (1958). Thus, any construction process or critique of an argument must be orchestrated by these elements. However, reasoning must be understood as a broader aspect than argumentation, as the latter is a specific way of reasoning (Walton, 1990).

In science education, Kuhn's work (1992, 1993) has been pioneering on learning through argumentation as well as studies on discursive processes in the classroom, where learning is understood through the analysis of communication or classroom discourse (Cazden, 1991; Hennessey, 1991; Sutton, 1992, 1997).

According to Revel et al. (2005), scientific argument in school is defined as "the production of a text in which a natural phenomenon is subsumed under a theoretical model by a mechanism of analogous nature" (p. 2). These authors identify similarities between the theoretical model, which is abstract or epitomic, and its counterpart, which is the theoretical reconstruction of the event. They also indicate that establishing this similarity does not imply the syntax of the argument to be an analogical reasoning, as this syntax may involve different types of strategies (deductive, inductive, abductive, etc.).

In the present study both erudite and school scientific argumentation are understood as a procedure, which is comparable to the practical skills or the cognitive and communication skills that are needed to produce, evaluate and apply science (Newton, Driver, & Osborne, 1999; Oliva, Aragón, & Cuesta, 2015). Sanmartí (2003) believes that this procedure has a cognitive-linguistic nature because it relies on highly complex cognitive skills and, at the same time, conveys itself through oral or written language in texts.

The capacity for argumentation relates explanations and evidence, with the latter employed to evaluate statements, theories or models (Jiménez-Aleixandre, & Puig, 2010). In this sense, the importance of evidence is directly related to scientific expertise suggested by different approaches, e.g., interpreting data and scientific evidence from the Programme for International Students Assessment (PISA) (OECD, 2013a), the ability to interpret results within a dimension of scientific competences called data analysis, or drafting conclusions according to Franco-Mariscal (2015) who approaches to inquiry-based science learning.

**Toulmin's Epistemological Model**

From an epistemological point of view, Toulmin's model is most commonly used when studying argumentation in science education (Erduran, Simon, & Osborne, 2004; Henao, & Stipcich, 2008; Pinochet, 2015). According to Toulmin (1958), science involves culture in constant transformation, as it poses questions and problems or involves explanations. From this point of view, learning science involves appropriating the cultural heritage, sharing the meanings and, at the same time, having the ability to take critical positions and change them. Toulmin (1958) considers the evolution of science as a plural, dynamic and community process of interaction among explanatory theories, in which argumentation – as the outsourcing of substantive reasoning – becomes the expression of a local and contingent rationality that allows changes to occur (Toulmin, 1958, 2003). This

author also suggests that the content of a certain science involves a set of established explanatory procedures and a number of tentative conceptual variants. The evolutionary change in these variants is governed by a general consensus on the criteria for the selection of variants and explanatory ideals for these processes (Toulmin, 1972).

Beyond a heuristic procedure and an analytical strategy, Toulmin's model on argumentation suggests a process that allows for social construction and negotiation of meaning, since it involves a dynamic dialogue in which we must – in order to sustain a statement, conclusion or point of view – provide reasons, receive cross-questions on the relevance of these reasons, confront objections and, perhaps, change or clarify an initial statement or point of view (Toulmin, Rieke, & Janik, 1979).

**Strategies to Develop and Assess Argumentation**

There are different proposals on how to approach argumentation in class (Kind, Kind, Hofstein, & Wilson, 2011; Sampson, & Blanchard, 2012). Felton and Kuhn (2001) propose that the quality of argumentation is different depending on whether it involves young people or adults, with the process being more elaborate and complex in the latter case. Therefore, both primary science students and PESTs should understand how to work and improve the argumentation competence in order to justify the training time invested in this area (Kuhn, 2010).

From an educational perspective, apart from having evidence of work, some researchers have attempted to develop new methods for improving the argumentation competence in PESTs (Demircioğlu, & Uçar, 2012; Tuysuz, Demirel, & Yildirim, 2013; Iordanou, & Constantinou, 2014). According to Tüysüz, Demirel, & Yildirim (2013), PESTs' understanding of science can be improved in science class by solving problems in which students are required to provide arguments. Jonassen (2004) indicates that to assess adequately a skill as complex as problem-solving, more than one kind of assessment is required and suggests three aspects to consider (performance assessment, component skills, and argumentation). Cho and Jonassen (2002) agree that an important skill in problem-solving, especially with ill-structured problems, is the production of coherent arguments to justify solutions and actions. Their study showed that the provision of a constraint-based argumentation scaffold during group problem-solving activities increased the generation of coherent arguments. Moreover, Cetin

(2014) believes that students should be given instructions on argumentation in order to improve learning of new content in science.

In respect to the above, it is important to seek the participation of PESTs as they are in direct contact with students in education from an early age, when it is easier to shape thinking and scientific logic, and hence contribute to a change in the way of arguing in society (Cetin, 2014).

Another important aspect is the context of argumentation, as there are very different contexts of a theoretical or empirical nature, different explanatory models, decision-making, and confirmation of predictions or critical evaluation of conclusions. An interesting context is that of models that help to understand causal explanations. In casual explanations, the knowledge of the matter is submitted to review and the use of evidence interacts in order to assess knowledge – i.e., the skill of using evidence – and science learning – i.e., scientific literacy (Jiménez-Aleixandre, & Puig, 2010).

In this area, activities and strategies must be designed in the appropriate contexts to ensure that students understand and use the scientific discourse and models, while they handle situations related to science with critical thinking in order to be able to solve problems and establish a debate on real issues (Jiménez-Aleixandre et al., 2000; Jiménez-Aleixandre, 2002), in other words to make the topic interesting and significant for students. In Visser's work (2007) it was highlighted that there has been a decrease in the number of people who prepare for careers in science and it is necessary to encourage a new generation to engage with science. In this sense, it is worth stressing socio-scientific issues in argumentation problems in order to solve this issue (Simonneaux, 2001; Zohar, & Nemet, 2002; Sadler, & Zeidler, 2005). According to Martínez and Ibáñez (2006), these types of contexts and problems enable students to be personally involved in the activity, thus monitoring and assessing their own knowledge and feeling personal satisfaction when solving a given problem. This in turn enhances positive attitudes towards science and science learning because, if science is taught this way, it becomes relevant for people's lives and promotes students' motivation to approach this topic. Finally, experts consider role-playing to be a very appropriate activity to promote argumentation on socio-scientific problems (Simonneaux, 2001; Nielsen, 2012).

The activities suggested to students in the PISA Assessment Programme (OECD, 2016) provide opportunities to address these contexts and develop different skills in young

people (Meyer, & Benavo, 2013). The argumentation competence is found among the scientific competencies that are assessed in the programme (Bybee, & McCrae, 2011). In order to improve the PISA score, Tsai (2015) proved that scientific competencies could be improved by using online argumentation. His study took PISA activities as a starting point to use questions to measure the capacity for argumentation, as the ultimate goal is for future teachers to be prepared to develop these skills in students.

In summary, for PESTs to learn how to design and assess argumentative activities, it is important to provide them with a simple argumentation model that enables them to learn and implement it easily with their students. The model used in this research is a simplification of Toulmin's model suggested by Jiménez-Aleixandre (2010) to facilitate the understanding of the essential elements that a good argument must have: namely evidence, justification and conclusion. 'Evidence' (aka *facts* in Toulmin's model) is the necessary proof to support a claim in a particular topic. 'Justification' (aka *warrant* in Toulmin's model) is the rationale behind the relationship between the evidence and the claim. 'Conclusion' (aka *claim* in Toulmin's model) is the opinion on a particular aspect. PESTs are required to identify these three elements in an argument and use them to design rubrics to assess arguments. Once this objective has been achieved, the model may consider further elements, such as 'previous ideas' or 'backing ideas', and 'rebuttals' or 'refutations', which are understood as limitations or weaknesses found in the conclusion and will serve as counter-responses.

Despite the unquestionable interest in improving competence with argumentation for both PESTs and students at all levels, the literature on this topic is scarce. Only a few studies have addressed the way in which to assess students' argumentative competence (Osborne et al., 2016), possibly because it is difficult to work with this competence in class and school laboratories (Hofstein, & Lunetta, 2004).

Different methods have been reported in the literature to assess scientific argumentation and some of these (Clark, & Sampson, 2008) are based on questionnaires that repetitively take into account the basic elements of Toulmin's model. However, other authors prefer to use instruments such as rubrics to escalate different levels of achievement of the argumentation competence (Osana, & Seymour, 2004; Jonassen, 2004; Özçinar, 2015; Deng, & Wang, 2017).

Assessing argumentation is not a simple task because different teachers may apply different criteria (Sampson, & Blanchard, 2012). For this reason teachers should choose a

good methodology. Different authors agree that, in order to understand and internalise the criteria for a good argument, students need to practice such criteria by assessing one another through peer assessment (Osborne et al., 2016). A number of research studies have proven that students learn to assess when they actually assess to learn (Boud, Cohen, & Sampson, 1999; Cebrián-Robles, Serrano-Angulo, & Cebrián-de-la-Serna, 2014). In this respect, Boud et al. (1999) made some recommendations for peer assessment in order to achieve greater success in student learning. The recommendations state that peer assessment must be conducted when the task involves extra motivation for students or when it is more helpful than it is problematic when compared to a more traditional methodology. Furthermore, peer assessment activities and tasks should be carefully designed in order not to 'devalue the evaluation', for instance, by evaluating numerous aspects in a single task. These recommendations were considered in the present study.

In general terms, formative assessment significantly improves students' scores at the end of the course (Black, Harrison, Lee, Marshall, & William, 2003). Other studies have highlighted that teamwork, where students can see and assess their peers' arguments, enables them to improve the quality of their arguments over time (Chin, & Osborne, 2010; Evagorou, & Osborne, 2013).

In summary, the importance of learning to argue involves teachers knowing how their students form their arguments. More specifically, they should know what their students consider as evidence, justification or conclusion, what rationale they use and when, in which situations they use models, concepts, laws or principles to justify their conclusions, what processes they use to argue while solving problems or epistemic levels that can be identified in their arguments (Kelly, & Takao, 2002; Jiménez-Aleixandre, & Díaz, 2003; Custodio, & Sanmartí, 2005). All of the above are key aspects to reason and argue, hence our interest in suggesting and evaluating strategies to train PESTs to improve the performance of their students in relation to their argumentation competence.

## Purpose of the Research

This study concerns the design and implementation of a training programme that integrates a set of strategies aimed at teaching Spanish PESTs how to argue. Secondly, the potential impact of the programme in relation to argumentation competence and its transfer to

educational practice was analysed. In order to achieve the goals the following research questions were posed:

Research Question 1: What level of argumentation competence do PESTs have before and after completing the training programme on argumentation?

Research Question 2: Are there differences in the argumentation competence shown by PESTs who have undergone this programme when compared to those who have followed the usual course syllabus with a more traditional teaching methodology?

Research Question 3: To what extent are the PESTs able to transfer to practice the knowledge that they have learned in this training programme?

## Methodology

### Sample

A total of 98 PESTs participated in this study. The PESTs' ages ranged from 20−21 and they were third-year students from the Primary Education Teaching Degree at the University of Malaga (Malaga, Spain). Students belonged to two samples. The first sample (experimental group) attended the training programme on argumentation, while the second sample (control group) followed the usual course syllabus, which did not include specific activities on argumentation.

The experimental group was composed of 57 students (E-PESTs). A total of 24.6% of the participants were male and 75.4% were female. The participants were randomly selected to undergo a training programme on argumentation as part of their 'Science Teaching' course syllabus, which was taught from March to June 2016. E-PESTs were taught by the first author of this paper, who has extensive experience in argumentation and who also served as a researcher and observer at the same time. The subjects for the control group were 41 students (C-PESTs) with similar characteristics to the experimental group. 39.0% of the participants were male and 61.0% were female. C-PESTs were taught by the second author of this paper. The participating E-PESTs and C-PESTs had not previously been trained in argumentation. In the results section, the direct quotations taken from the responses of PESTs are identified by the corresponding group abbreviation (E - experimental group or C - control group) followed by the number used to code that particular PEST.

The ‘Science Teaching’ course syllabus at the University of Malaga included modules on the design of science teaching units, selection and presentation of scientific contents, experimentation in class, scientific attitudes and principles, and evaluation. However, the course did not include any specific module on argumentation and hence the training programme for E-PESTs. Most of the PESTs in this study only had knowledge of science education corresponding to K-12 level.

**Training Programme on Argumentation**

The experimental group participated in a training programme of six sessions (each 120 minutes long) that addressed theoretical aspects of argumentation, such as Toulmin’s simplified model (Jiménez-Aleixandre, 2010) and a set of specific activities on argumentation (Figure 1). On the one hand, the programme highlighted the theoretical nature of arguing compared to other processes like describing, defining, summarising or storytelling. On the other hand, it was intended that the PESTs would become aware of their future students’ difficulty in elaborating arguments.

In summary, the aim of the programme was to develop the competence of argumentation in PESTs, which will provide them with tools to elaborate arguments in the first instance and to be able to criticise and refute others’ ideas, thus improving their critical thinking about science in the second place (Walton, 1989; Driver et al., 2000). In order to achieve these goals, the programme emphasised motivation and the importance of educating argumentation skills from a very early age, i.e., in Primary Education students. As a result, new generations will eventually be prepared to receive, judge, and create scientific information.

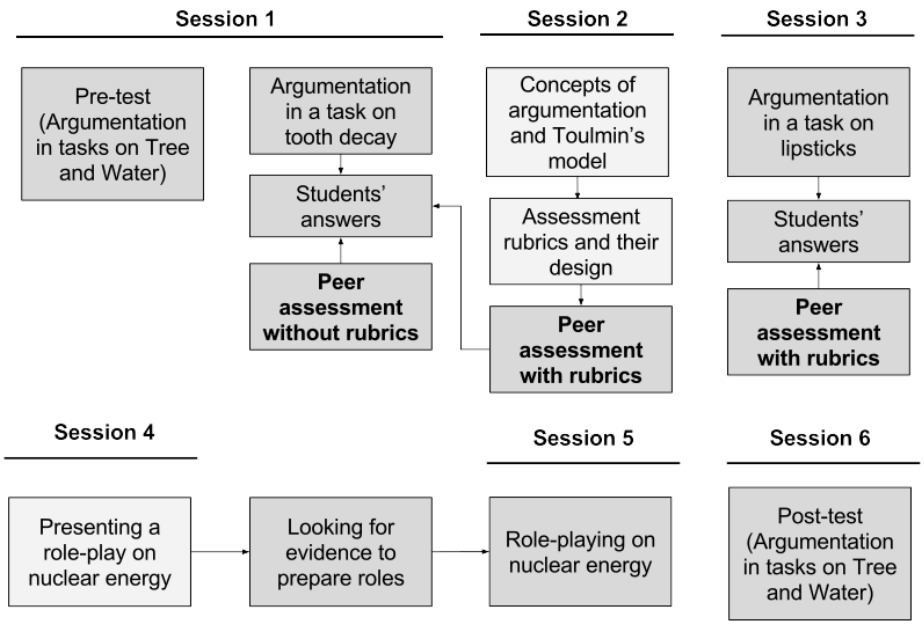


**Fig. 1** Outline of the training proposal to study argumentation in class

This training programme focused on the active role of PESTs as they were required to conduct a number of tasks related to argumentation. To this end, the programme included the following strategies to help improve PESTs' argumentative competence: (a) identification of the essential elements in a written argument; (b) designing evaluation rubrics based on a basic rubric; (c) peer assessment with and without rubrics; and (d) role-playing to promote argumentation on socio-scientific issues (Simonneaux, 2001; Nielsen, 2012). The aspects are shown with a darkened background in Figure 1.

Based on the approach described by Osborne et al. (2016) for argumentation competence, the training programme addressed a number of levels of learning progression (p. 8). The programme began and ended by asking PESTs to develop a complete argument (level 1c) in two different contexts at the beginning, as an indicator of their initial level of competence with argumentation. When students assess their peers' arguments using rubrics, they have to prove their ability to identify the essential elements in a good argument (0b, 0d and 1d levels). The preparation and development of role-playing demand seeking evidence (0c level), elaborating arguments (1c level) and criticising other students' arguments (2a level) and these tasks are subsequently carried out.

The new perspective of this proposal is that it combines peer assessment with specific activities on argumentation. These aspects have already proven their effectiveness in other research studies. Boud et al. (1999) showed how peer assessment improves student learning. All of this creates a rich educational framework in which PESTs were offered different opportunities, through contexts and contents, to enhance their argumentation competence.

The training sessions of the programme were included in course modules in order to relate their content with the content of the course syllabus. The first and sixth sessions were conducted at the beginning and the end of the course, as a pre-test and post-test, in order to observe the evolution of PESTs' argumentation competence. These sessions are described below.

*Session 1*

This session had two parts. The aim of the first part was to evaluate PESTs' initial argumentation competence. PESTs had to answer two scientific questions (Annex I), one related to biology and one to chemistry. These activities were adapted from questions designed in various educational assessment documents in line with the intellectual level required in PISA tests (OECD, 2016).

The first activity focused on tree growth and how trees form rings in their trunks as they grow and this was suggested in the general diagnosis assessment 2009 for Primary Education from the Spanish Ministry of Education (Ministerio de Educación, 2010). The test showed a correlation of 0.761 with a similar task included in PISA 2006 in the physical section and a reliability of 0.835. The second activity involved chemical concepts within an everyday context: allusions to the purity of bottled water in advertising. This question was taken from a research study conducted with 15-year old students (Rodríguez-Mora, 2016).

Both questions can be considered equivalent in some respects because they include a text explanation and a drawing, and they contain the evidence necessary to reach the conclusion and provide the justification. The two questions were formulated and used initially with students of similar ages (15 years). The fundamental difference between the two, in addition to the context to which they relate, lies in the conceptual demand. A great depth of scientific knowledge was not required for the PESTs to argue about the tree growth activity. However, the students required knowledge of the concepts chemical solution and

pure substance to argue about the purity of bottled water. The literature shows how students have problems in differenting these two concepts adequately (Furió, & Domínguez, 2007).

The aim of the second part of the session was to ascertain the PESTs' initial capacity to assess argumentation. To do this, they were asked to answer an argumentative question drawn from PISA (OECD, 2006a) on tooth decay, which addressed the influence of bacteria in the formation of tooth decay (Annex II). Once answered, PESTs were asked to assess their randomly chosen peers' answers. Each PEST then became responsible for assessing their peer's answer and was required to justify the score given. PESTs did not know the peer's name and were not allowed to ask for help or use pre-established assessment criteria to evaluate rubrics. PESTs assessed using their own criteria and rated answers on a 0–10 scale, providing justification for their assigned value.

*Session 2*

This session addressed theoretical aspects of the appropriate contexts to argue, Toulmin's model, essential elements of arguments and the presentation of a basic rubric to assess any argumentative activity in general terms (Figure 2) based on which specific rubrics can be designed.

The tooth decay example from Session 1 was used to explain the argumentation model. The question as to why the context of oral hygiene is of interest to primary education students was explained to PESTs first. Specifically, because it poses a real problem at their age – one in three children have tooth decay (Llodra, 2012) – and can become a social problem in secondary education, where students may be dissatisfied with their appearance because of tooth decay (Sheiham, 2005). The adaptation from the PISA test (OECD, 2006a) was also explained. This consists of using the exact same question as in the PISA test, only adapting it to motivate students towards a reasoned response. The elements of the argument were identified and used to build a specific rubric for assessment, which was drawn from the basic rubric.

| 1.Evidence | | | |
|---|---|---|---|
| 1 There is no evidence | 2 There is some evidence, but inappropriate to support the conclusion | 3 There is some evidence but not sufficient. Some evidence given might be inappropriate | 4 There is enough evidence, and it is appropriate to support the conclusion |

| 2.Justification | | | |
|---|---|---|---|
| 1 There is no justification | 2 There is a justification, but it does not link the evidence with the conclusion | 3 There is a justification that links the evidence with the conclusion. It repeats the evidence and/or some scientific ideas are included but they are insufficient | 4 There is a justification that links the evidence with the conclusion. It includes the appropriate and sufficient scientific ideas |

| 3.Conclusion/s | | | |
|---|---|---|---|
| 1 There is/are no conclusions | 2 The conclusion is not right | 3 The conclusion is right, but it bears scientific mistakes | 4 The conclusion is right and bears no scientific mistakes |

**Fig. 2** General basic rubric to assess arguments (source: the authors of this research)

The basic rubric used (Figure 2) has three indicators to evaluate the three main elements in an argument: evidence, justification and conclusions. Each indicator is presented in four levels of success or achievement, corresponding to levels 1 to 4, where 1 is the lowest and 4 the highest degree of accuracy.

This study uses “flexible” rubrics instead of “classical” or “squared” rubrics. In this approach the achievement levels are not fixed in each rubric element (as would be expected in a traditional rubric), so as to allow for freedom and objectivity in the rubric design (Cebrián-de-la-Serna, & Monedero-Moya, 2014). In this way, PESTs can use the base rubric as the starting point to build as many specific rubrics as argumentative activities are conducted in the classroom. When designing rubrics, the number of achievement levels can vary and the qualitative text of each level must adapt to the activity in question.

The specific rubric on tooth decay is designed with a different number of achievement levels from the basic rubric in order to reflect whether the question is asked in a restricted or tentative way. Once the model is explained along with how to evaluate with rubrics, PESTs

conducted a second peer assessment with their answers, but this time using the specific rubric. The aim was to check whether there was any more consistency in the results when using the specific rubric.

The second assessment was conducted using an e-rubric (CoRubric) (Cebrián-Robles, 2016) designed by one of the authors of this research – this e-rubric is available free of charge. The advantages of the CoRubric are as follows: peer assessment (formative assessment), which allows for multiple evaluations to check progress, role management using notes and comments and exporting results to other programmes for treatment. Thus, students can assess and justify their scores through online rubrics, where they can conduct a 360º assessment, i.e., peer assessment, self-assessment and teacher assessment of the questions. Finally, students should identify their formative evolution regarding their capacity for argumentation by answering these questions at different times.

*Session 3*

In this session, PESTs answered a new argumentation question to assess whether they had really improved their argumentative quality. The question addressed the hardness of two brands of lipstick and how they could soften the composition by changing the ingredients. The heading was extracted and adapted from PISA 2006 science tests (OECD, 2006b) (Annex III). To conclude this task, PESTs were asked to conduct peer assessment with at least two assessments per student and using a specific rubric provided by the teacher.

*Session 4*

The fourth session of the training programme presented and explained the nature of role-playing and why it is of interest in Primary Education, as well as the rules to play the game in Session 5. PESTs were then assigned roles and asked to seek and identify evidence that would allow them to defend their role, together with counter-arguments that would allow them to respond to potential rebuttals from their peers. They were given a week to do this, before Session 5, and it was estimated that they would need to dedicate 8 hours to this task. Their role must not be disclosed to any of their peers, not even outside class, as one of the goals of the game was to relate each PEST to their role based on evidence.

The role-play concerned a current socio-scientific issue: the nuclear power plant in Garoña (Burgos, Spain). The task was to take a decision about whether to extend the permission to keep the plant running for 20 to 40 years or not. In order to take such decision,

PESTs were randomly given different roles to play (shareholder, businessperson, lobbyist, villager, scientist, etc.). Besides delving further into the issue of nuclear energy, PESTs were shown an advanced aspect of argumentation, i.e., refutation of ideas to improve the quality of arguments. One of the characters was a minister who acted impartially and must be convinced by the quality of arguments in order to make the right decision. Thus, PESTs were motivated to prepare themselves very well, to win the game by convincing not only those who were in the opposite position but also the minister, who acted as the jury and ultimately decided who won.

*Session 5*

The role-play was carried out in Session 5, for which PESTs had been preparing the week before. Every PEST argued their position without revealing the role they had been given. Meanwhile, the PEST who played the judge (minister) took notes of the different views and arguments and decided which of the two groups was stronger, and the reason why the permission for the nuclear plant must or must not be extended. All judges (from all role plays) presented and reasoned their final decisions at the end of the session.

*Session 6*

The sixth session was held on the last day of class and PESTs were asked to do a post-test on the two questions (tree growth and purity of bottled water). The aim was to check for differences in the quality of their arguments after completing the training programme. PESTs were not given any previous information about the answers to these two questions.

**School Practices of PESTs**

The E- and C-PESTs who participated in this research carried out 13 weeks of school practice in an elementary school, which began four months after the end of the proposed training programme. In these practices, the PESTs had to design and implement a teaching unit with students' aged from 6 to 12 years. At the end of the period the PESTs produced a portfolio in which this teaching unit was included.

These teaching units are not predefined in the training programmes of their degree and could include content from any subject. Thus, the PESTs were the ones who decided the specific work topic, the content and their approach to teaching activities, but the content was always agreed with their academic and professional tutors, who were not involved in the

research presented here. Therefore, PESTs could freely decide whether or not to include argumentation activities as part of the teaching unit.

The PESTs submitted the practice portfolios 6 months after completing the training programme. For this project it was only possible to access the portfolios that were deposited in the official repository established by the faculty. In this way, 25 portfolios from E-PESTs and 29 from C-PESTs were accessed.

**Data Analysis**

Analysis of the data was carried out in two phases. The first phase corresponded to the analysis of the students' responses in the pre- and post-test of the training programme. The second phase addressed the analysis of the argumentation found in the PESTs practice portfolios.

**Data analysis pre- and post-test in the training programme**

The procedure used for data analysis in the pre- and post-tests included the following steps: (1) answering the questions raised correctly; (2) using these answers to design a specific rubric to assess each activity; and (3) identifying the elements of arguments in PESTs' responses and assessing them with the designed rubric.

*1. Answering the Questions Raised Correctly*

Firstly, the most appropriate answers to the questions raised in the pre- and post-tests (Annex 1) were established and the argumentation elements observed in them were identified. To do so, the authors of this research agreed by consensus on the correct answers, taking into account the views of a number of experts on the questions, at the highest level in the rubrics. The most appropriate answers to both questions are shown in Table 1 and these are considered the highest achievement level in the specific rubric designed for assessing these questions.

**Table 1** Appropriate answers for each element in the argument

| | Tree Growth Activity | Purity of Bottled Water Activity |
|---|---|---|
| Appropriate answer | In the stem of the plant, a well-defined ring represents a year of life. The width of the ring depends on the weather conditions of the year and is higher when these conditions are favorable to | Bottled water is not chemically pure since it contains different solutes, as you can see on the label, which indicates the presence of gas, calcium and sodium. Therefore, bottled water is a mixture or |

| | | |
|---|---|---|
| | the tree. Therefore, the X ring is the one with the greatest width and the most favorable weather conditions. | solution. Pure water is $H_2O$. |
| Evidence | E1. The widest ring is the one that has grown the most (that year). | E1. The label indicates that it is sparkling water. |
| | E2. Therefore, the X ring is the one with the greatest width. | E2. It has calcium and sodium in its composition. |
| Justification | J1. The X ring is the one with the best weather conditions. | J1. It is not chemically pure because it contains different solutes.<br>J2. Pure water contains only water.<br>J3. Bottled water is a mixture (solution). |
| Conclusion | The width of the ring corresponds to the amount of matter synthesized by the tree during the year. Tree growth is favoured by the right weather conditions (rain, sun, etc.). | Bottled water is not chemically pure. |

*2. Design of Assessment Rubrics*

Specific rubrics were designed based on the most appropriate answer and adapted to each activity. For example, the rubric designed to assess the purity of bottled water activity is shown in Figure 3. In this rubric there are four achievement levels for evidence and five for justifications and conclusions. Level 1 is the least desirable and level 4 or 5 (depending on the element) is the best in terms of argumentative quality.

| 1.Evidence | | | | |
|---|---|---|---|---|
| 1 There is no evidence | 2 Although there is evidence, it has nothing to do with water purity | 3 It provides some evidence that justifies the reason the water is not pure from a scientific point of view | 4 Provides enough and appropriate evidence to defend that the water is not pure from a scientific point of view (which has gas, calcium and sodium) | |

| 2.Justification | | | | |
|---|---|---|---|---|
| 1 There is no justification | 2 There is a justification, but it does not link the evidence with the conclusions or does so without chemical knowledge | 3 There is a justification that links the evidence with the conclusion/s but it is not enough to argue that the water is not pure from a scientific point of view and/or lacks chemical knowledge | 4 Provides a justification that links the conclusion with the evidence. There are sufficient correct scientific ideas with implicit chemical knowledge, but the terms are not used accurately or are done so tentatively | 5 Provides a justification that links the conclusion to the evidence. There are sufficient correct scientific ideas with implicit chemical knowledge. Terms are used accurately and specifically |

| 3.Conclusion/s | | | | |
|---|---|---|---|---|
| 1 There is/are no conclusions (Unknown or it says it cannot be known) | 2 Reaches an erroneous conclusion by thinking that the water is pure scientifically | 3 Despite it being the right answer, it has not been well defended scientifically and/or it mixes the colloquial meaning of "pure" with its chemical meaning | 4 The conclusion is correct and well-defended scientifically but uses tentative terms (I think, in my opinion...) | 5 The conclusion is correct and well-defended scientifically and specifically |

**Fig. 3** Assessment rubric for the activity on the purity of bottled water

While the most interesting aspect of this rubric is to ascertain the quality level of achievement for each element in an argument, an overall score was also awarded to provide an idea of the overall quality of the argument. The score consists of a quantitative variable from 0 to 100 points. To this end, each indicator (evidence, justification and conclusion) has the same value, so that the intellectual leap between the different levels of achievement is the same. Thus, in the case of justification and conclusion each intellectual leap is worth 1/4 as it has five levels of achievement, while in evidence the value of each leap is 1/3 as it has only four levels.

The final overall performance of PESTs in relation to their argumentative quality was obtained by adding individual scores for each indicator and multiplying this number by the percentage of each indicator (33.33%). For instance, an argument whose evidence is at level 2, justification at level 3 and conclusion at level 4, would obtain an argumentative quality as follows: $33.33 \times 0.33 + 50.00 \times 0.33 + 75.00 \times 0.33$, which would result in 52.78 out of 100 points. A good argument involves an achievement level of 3 in each element of the rubric.

3. *Identification of the elements of arguments in PESTs' answers and assessment with the designed rubric*

PESTs' answers were analysed and the elements of arguments were identified in their answers. This was done individually by the three researchers, who combined their analyses and reached a consensus in cases where there was no agreement. In the example below, two answers of an E-PEST were identified and analysed in the pre- and post-tests in relation to the purity of bottled water activity:

> "*No, truly pure water is distilled water*" (E-PEST14, pre-test, purity of bottled water activity)
>
> *"This bottled water cannot be pure because we can see in the specifications that it has 0.06 of calcium and 11.3 of sodium. This means that water has other components apart from $H_2O$ (water). For water to be completely pure it should only have $H_2O$ molecules"* (E-PEST14, post-test, purity of bottled water activity)

In the pre-test answer, E-PEST14 was only able to use two out of three elements of a good argument: justification and conclusion. As justification, this E-PEST stated that "*truly pure water is distilled water*", in relation to justification J2 (Table 1). Clearly, this answer lacked evidence and the given justification was not entirely true, but the conclusion was correct.

On using the rubric in Figure 3 to assess this PEST's answer, his/her conclusion (solid line) was at level 5, as argumentation here reflects that the concept of purity is correct, even though the sentence is short. However, evidence is at level 1 because it wasn't provided and justification (dashed line) is at level 2 because, although present, it lacks evidence to support the conclusion.

The answer given by this E-PEST after the training programme was much more elaborate as it reflected all the elements of an argument in a clear way. S/he indicated at least one out of two aspects of evidence (E2, Table 1) by mentioning that this water contains calcium and sodium, although s/he missed the gas. Nevertheless, the justification is right as it referred to E2 when reasoning that water can only have water in its composition. Finally, the conclusion was also correct as it introduced the concept of water purity and insisted that this water could not be pure.

Assessment rubrics show how E-PEST14 progressed in all elements of an argument and reached the highest levels of achievement in all cases except in evidence, where s/he missed one aspect. Therefore, after completing the training programme this PEST was at level 5 for conclusions (solid lines), as the given conclusion was reasoned in a scientific way; at level 3 for evidence (dashed lines and dots), as one out of two pieces of evidence was presented; and at level 5 for justification (dashed lines), as s/he restrictively indicated that water was not chemically pure because it contained various dissolved substances and pure water can only contain water.

**Data analysis for the argumentation found in the practice portfolio**

As a first step, a search was carried out for the possible references to the argumentation in the practice portfolios by reading each portfolio. In addition, portfolios were available in electronic format (PDF) and it was possible to do a search in Adobe Acrobat Reader to confirm that all of the references to argumentation had been analysed. In the search the keyword fragment "argu" was used allows to locate different words related to argumentation.

The second step addressed the analysis of the extent to which the argument was mentioned in the portfolios. This was achieved by establishing a system of progress that had three categories for the ideas learned during the training programme:

C1. Argumentation is not mentioned in the practice portfolio.

C2. Argumentation is mentioned, but it is not part of the teaching unit activities. This could be an indicator that students have considered argumentation as interesting in their training period, even though it was not specifically included in the activities of their teaching unit. An example is the use of argumentation as one of the aims of the teaching unit but without associating a defined activity: "*To listen, to speak and to dialogue in situations of communication in the classroom, arguing about their work, manifesting a receptive attitude and respecting the external approaches*" (E-PEST05).

C3. Argumentation is part of the teaching unit activities that were put into practice. This may indicate that the PESTs have considered that argumentation should be part of their work plan and it was also put into practice. The following examples of portfolios highlight how PESTs used argumentation as part of the teaching activities:

*"In order to assess the exercises, I always make the student argue how and why he/she did it that way, which in this case has been very useful, since many of the activities were problems or required logical reasoning"* (E-PEST20).

*"Participation was very active and we used the dialogue and argumentation as the basis for carrying out such activities"* (E-PEST03).

**Statistical Analysis**

Data were processed with the statistical software package SPSS 23.0. In order to check if the rubric variables and the total score represent a normal curve the Kolmogorov–Smirnov test was performed. The result showed 0.000 significance in all variables, thus rejecting the null hypothesis stating that there was normality in the variables. The test showed that scores and achievement levels in the results of the rubric variables were not normally distributed and, for this reason, the Mann–Whitney U and Wilcoxon tests were suggested.

The Mann–Whitney U test compared the starting point for PESTs to ascertain whether there were any significant differences between groups at the beginning of the process. The Mann–Whitney U test was also conducted in the final stage to ratify that the results obtained at the end of the training programme differed significantly in the experimental and control groups. The Wilcoxon test checked for differences before and after the training programme.

The effect size of the Wilcoxon and Mann–Whitney U statistical tests was calculated using the equation $r = Z / \sqrt{N}$, where $N$ is the number of PESTs over the two time points and $Z$ is the value of the statistical test. As for the value of $r$ calculated in absolute terms: 0.1 is considered a small effect, 0.3 a medium effect and 0.5 a large effect (Cohen, 1988).

The Fisher's exact test was used to verify the existence of significant differences between the E- and C-group in the transfer to practice of the argumentation. This test was used on the one hand to determine the differences when mentioning or not mentioning the argument and on the other hand to determine the degree of use of the argumentation in its transference.

## Results and Discussion

**Results of the pre- and post-test in the teaching programme**

In this section the results before and after the intervention (pre and post) are analysed. The PESTs' starting point was analysed first in order to check their level of argumentation

competence, i.e., their skills to offer a complete argument, and then to ascertain whether there were any significant differences between the groups. The PESTs' final status was then analysed by considering two perspectives. On the one hand the final status for each group and on the other hand differences between the experimental and control groups were analysed in the post-test to check if initial similarities had changed after the intervention. This analysis was used to check the possible influence of the training programme.

All of the results of the study are displayed in Table 2, which shows for each group (E and C) and activity the values of Median (Mdn) in each element of the argument and the overall performance in argumentation in the pre-test and post-test. The results of the Wilcoxon test (*Z* parameter, *p* probability and *r* effect size) and Mann–Whitney U test (*U* and *Z* parameters, *p* probability and *r* effect size) are also included.

**Table 2** Results of the study

| | Experimental Group | | | | | Control Group | | | | | Experimental vs Control Group | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Pre-test | Post-test | Wilcoxon Test | | | Pre-test | Post-test | Wilcoxon Test | | | Mann-Whitney U Test (pre-) | | | Effect Size | Mann-Whitney U Test (post-) | | | Effect Size |
| | Mdn | Mdn | Z | p | r | Mdn | Mdn | Z | p | r | U | Z | p | r | U | Z | p | r |
| Tree Growth Activity | | | | | | | | | | | | | | | | | | |
| Evidence | 3.0 | 4.0 | -3.3 | 0.0 | 0.31 | 3.0 | 3.0 | -1.0 | NS | -- | 973.0 | -1.52 | NS | -- | 1000.0 | -1.39 | NS | -- |
| Justification | 3.0 | 5.0 | -5.2 | 0.0 | 0.49 | 1.0 | 1.0 | -0.5 | NS | -- | 736.0 | -3.22 | 0.0 | 0.32 | 385.5 | -6.11 | 0.0 | 0.62 |
| Conclusion | 5.0 | 5.0 | -2.7 | 0.0 | 0.25 | 5.0 | 5.0 | -0.3 | NS | -- | 1146.5 | -0.18 | NS | -- | 1053.0 | -1.16 | NS | -- |
| Overall performance in argumentation | 66.7 | 100.0 | -5.1 | 0.0 | 0.48 | 58.3 | 66.7 | -0.7 | NS | -- | 1028.0 | -1.02 | NS | -- | 531.5 | -4.75 | 0.0 | 0.48 |
| Purity of Bottle Water Activity | | | | | | | | | | | | | | | | | | |
| Evidence | 3.0 | 3.0 | -4.5 | 0.0 | 0.42 | 3.0 | 3.0 | -1.1 | NS | -- | 1012.5 | -1.29 | NS | -- | 922.5 | -2.05 | 0.0 | 0.21 |
| Justification | 2.0 | 3.0 | -5.1 | 0.0 | 0.48 | 2.0 | 2.0 | -0.8 | NS | -- | 1091.5 | -0.64 | NS | -- | 526.0 | -4.81 | 0.0 | 0.48 |
| Conclusion | 3.0 | 4.0 | -3.4 | 0.0 | 0.32 | 3.0 | 3.0 | -2.2 | 0.03 | 0.29 | 1118.0 | -0.43 | NS | -- | 998.0 | -1.37 | NS | -- |
| Overall performance in argumentation | 47.2 | 66.7 | -5.5 | 0.0 | 0.52 | 47.2 | 47.2 | -1.1 | NS | -- | 1032.0 | -1.00 | NS | -- | 681.5 | -3.53 | 0.0 | 0.36 |

***Starting Point***

With regards to tree growth activity, the experimental group scored well in the pre-test (Table 2) on the three elements of arguments. The best results were found in the conclusion (4.2). Evidence and justification gave lower results than conclusion, with around level 3 obtained in each case. The control group showed a lower level in the quality of justifications in the pre-test, with only level 2 achieved.

The following responses illustrate these differences in the initial justification capacity between the PESTs in both groups. The justifications are shown in italics. Evidence and conclusion are indicated in small capitals and underlined text, respectively.

"<u>The ring with the letter "X</u>" since during that year there has been A GREATER WIDTH OF THE TRUNK RING. Therefore *it can be intuitively seen that in that period of time the climatic conditions were more favorable*" (E-PEST23, pre-test) (Evidence: level 3; Justification: level 3; Conclusion: level 5)

"<u>The X ring</u>, since IT IS THE THICKEST, and *its width is due to the good weather that has barely worn away the trunk*." (C-PEST18, pre-test) (Evidence: level 3; Justification: level 2; Conclusion: level 5).

The results for the purity of bottled water activity were less favorable in the pre-test in both groups (Table 2), with lower scores obtained than in the tree growth activity. The best results were once again observed in conclusion (3.5). These results can be explained in terms of the role that scientific knowledge may play in the argumentation competence (Yaman 2017). Presumably, this is because PESTs needed to know the chemical concept of purity of water before providing a coherent argument. Many of the participants referred to the concept of purity as a colloquial term: *"Yes, it is pure, as the label says it does not contain any added component"* (E-PEST48, pre-test) or understood purity as the water that comes from nature: *"It is impossible for pure water as found in nature to have those 0 levels, which means that it must have been modified"* (E-PEST43, pre-test).

The medians values for both questions for the overall performance in argumentation seemed to show that the experimental group (66.7 for the tree growth and 47.2 for purity of bottled water) did not argue better than the control group (58.3 and 47.2, respectively). Initial differences between the groups in pre-test were evaluated by conducting the Mann–Whitney U test on the groups (Table 2). The results suggest that the experimental and control group

comparison did not show significant differences in their initial argumentative competence. However, significant differences were found in relation to justifications in the tree growth activity (U = 736.0; p = 0.00). The value of the effect size produced in the Mann–Whitney U test for this element were also calculated (r = 0.32) and the results indicated a medium effect size on using Cohen (1988) criteria.

***Final Status***

*Pre-test and Post-test Comparison for Each Group*

In this section it was determined whether the training programme managed to significantly improve argumentative competency in PESTs. The Wilcoxon test showed significant differences between pre-test and post-test in the experimental group (Table 2) in the tree growth activity in the three elements of an argument and in the overall performance. However, these differences were not observed in the control group.

Significant differences were also observed for the water purity activity in the same test, where results of the experimental group are similar but significant differences were found in conclusion for the control group. The reason for this is probably that the conclusion in water purity task was difficult to understand at the beginning, so some PESTs may have written their conclusions after the pre-test.

The improvement achieved by the E-PEST can be clearly seen in the answers given by the same student in the pre- and post-test:

"I think it is not really pure water *since it always contains remnants of substances*" (E-PEST24, pre-test) (Evidence: level 1; Justification: level 2; Conclusion: level 3).

"It is not a pure water, nor truly pure, *since for its bottling it has undergone numerous processes. The water also HAS A 0.06 OF CALCIUM AND 11.3 OF SODIUM. Therefore, it is not pure as it contains these concentrations of calcium and sodium*. ALSO IT IS WATER WITH GAS, and for this reason, we can say directly that it is not pure" (E-PEST24, post-test) (Evidence: level 4; Justification: level 3; Conclusion: level 5)

We can observe in the example how this E-PEST improves in the three elements of the argument and ultimately achieves the highest levels for evidence and conclusion, and a minor advance in justification.

The values of the effect size produced in the Wilcoxon test for the elements of the arguments were calculated. The r values were 0.25 and 0.50 for the experimental group. Although in general a large effect can be considered as the r value is very close to 0.5, some differences can be observed in each particular element. So, conclusion was the element with the least marked effect in both questions (r = 0.25 for tree growth and r = 0.32 for purity of bottled water activity), while justifications showed the greatest effect (r = 0.49 in both activities).

Finally, the results of the Wilcoxon test for the overall performance in argumentation showed significant differences for the experimental group in both activities, but not for the control group. Furthermore, the quality of the arguments in the control group hardly varied, with values similar to those obtained in the pre-test by the experimental group.

*Post-test Comparison between Groups*

Comparison of the final status between the experimental and control groups using the Mann–Whitney U test showed significant differences in justification but no differences for conclusion for both activities. Only in the water purity activity were differences found in evidence (Table 2).

In the following example the arguments given by a C-PEST and another E-PEST in the tree growth activity are compared. As can be seen, although the levels of evidence are similar, a better level of achievement for justification was offered by the E-PEST, with both PESTs reaching the same level of conclusion:

"The X ring was the year where the weather conditions were more favorable, since it is the one that HAS A GREATER THICKNESS" (C-PEST18, post-test) (Evidence: level 3; Justification: level 1; Conclusion: level 5).

"The tree ring that corresponds to the year in which the climatic conditions were most favorable is the X. As we know that THE BIGGER THE TREE GROWS THE WIDER ITS RING IS IN THAT YEAR, we know that the year corresponding to *the X ring had to have climatic conditions that were suitable for the tree and made its growth will not stagnate and grow more than other years*" (E-PEST31, post-test) (Evidence: level 4; Justification: level 5; Conclusion: level 5).

The following examples show how the argument of an E-PEST presents better evidence and justification than that of a C-PEST in the purity of bottled water activity, although both participants reach the same level of conclusion:

"It is not pure, since the pure water does not CONTAIN GAS, AS WE CAN SEE ON THE LABEL" (C-PEST039, post-test) (Evidence: level 3; Justification: level 1; Conclusion: level 5).

"Firstly, GAS HAS BEEN ADDED INTO THE WATER, so it can not be pure. In addition, *the water should only contain hydrogen and oxygen* to be "truly pure" and CONCENTRATIONS OF CALCIUM AND SODIUM ARE SHOWN ON THE LABEL" (E-PEST05, post-test) (Evidence: level 4; Justification: level 5; Conclusion: level 5).

In terms of the overall performance, significant differences were found for both questions. The effect size produced in the Mann–Whitney U test was calculated. The data obtained (Table 2) indicate that, in both activities, a medium effect size was obtained for the overall performance in argumentation.

Taking into account both activities, as for the elements of a good argument, it is worth noting that the conclusions did not differ significantly between the groups. This finding is because questions were simple enough for all answers to be correct and this is also why the control group did not differ from the experimental group in their conclusions.

The only significant differences were observed for the purity of bottled water activity with an effect size $r = 0.21$, which should be considered low for evidence. The lack of significance in evidence for the tree growth activity can be explained by the fact that PESTs initially showed high levels of achievement (Mdn = 4.0 and Mdn = 3.0 for E- and C- group, respectively), which were already close to the maximum level used in the rubric (4).

In the case of justifications there were significant differences in both activities, with a high effect size for the tree growth activity and a medium effect for the purity of bottled water activity.

Finally, it should be noted that the context of the activities used in pre- and post-test seem to influence the initial and final argumentation competence of the PESTs. The fact that better results were obtained in the tree growth activity than in the purity of bottled water activity in pre-test, and that the final differences between experimental and control groups were more pronounced in the purity of bottled water activity, could be explained in terms of the conceptual demands of the two activities. In this regard, it needs to be borne in mind that

scientific knowledge is an important component of argumentation competence (Osborne et al., 2016).

### Results of transference to practice

It is considered in this work that transference to practice occurs when the PESTs are able to include, either in the teaching unit design (C2 category) or in the activities carried out during the period of practice (C3 category), the techniques learned in the training programme, i.e., the different ideas about argumentation. The results obtained when analyzing the practice portfolios from this approach are shown in Table 3.

**Table 3** Results of transference to practice of argumentation

| | Experimental Group (N = 25) | | Control Group (N = 29) | |
|---|---|---|---|---|
| Inside the practice portfolio: | f | % | f | % |
| C1. Argumentation is not mentioned | 10 | 40 | 20 | 69 |
| C2. Argumentation is mentioned, but it is not part of the activities of the teaching unit | 2 | 8 | 5 | 17.2 |
| C3. Argumentation is part of the activities of the teaching unit which were put into practice | 13 | 52 | 4 | 13.8 |

The Fisher's exact test was used to analyze the possible differences between the E-group and C-group because of the nature of the data and the small size of the sample. There are significant differences between the three categories and the two groups ($p = 0.0098$; $p < 0.05$). This test was performed to ascertain whether argumentation was mentioned by PESTs in their practice portfolios (C1 vs C2 plus C3) and this showed no significant differences ($p = 0.0540$; $p > 0.05$) between groups. In addition, significant differences were found between the E-group and C-group regarding the degree to which argumentation was part of the activities (C1 plus C2 vs C3) ($p= 0.0035$; $p < 0.05$). Thus, the E-PESTs were more able to include argumentation as part of the activities in their teaching units that they put into practice.

## Conclusions and Implications

The results of this study allow a set of conclusions to be drawn in response to the research questions. In general terms, the training programme significantly improved the level of argumentation competence of PESTs in everyday situations (Klein, 2004). This finding is

demonstrated by the significant differences between pre- and post-test results of experimental groups in the three elements of arguments (evidence, justification and conclusion) (Research Question 1) and also by the even greater progress in some of the elements (justification in both activities and evidence in the purity of bottled water activity) and the overall performance in argumentation in both activities compared to the results obtained by C-PESTs (Research Question 2).

As shown in this study, the starting point for all PESTs regarding argumentation competence proved to have limitations when providing justification and evidence as stated in the literature (Andrews, & Mitchell, 2001; Osborne et al., 2016). The progress in the quality of PESTs' arguments was demonstrated by the improvement in identifying evidence and using it to justify conclusions and also in the quality of their justifications and conclusions as it is also mentioned in the literature (Demircioğlu, & Uçar, 2012).

The conclusions show that it is possible to improve the initial training in PEST argumentation by providing training programmes with specific argumentation activities in which PESTs are involved in peer evaluation using rubrics. Some of these activities need a close reading and cross-textual analysis in order to make the building blocks of argumentation suggested by Litman et al. (2017). However, the results also highlight the need to pay more attention to the scientific knowledge involved in the tasks of argumentation (Osborne et al., 2016). Likewise, the training programme should also have more impact on helping PESTs to identify evidence and to use it as an integral part of their arguments.

Despite the progress in the quality of PESTs' arguments, it is difficult to identify which elements of the training programme are directly responsible for this progress. However, getting PESTs involved in assessing their own learning, assessing that of their peers and designing assessment instruments all seem to have been key factors in this success.

Likewise, it should be noted that this training programme does not cover all of the elements that can be covered in relation to argumentation competence. However, the programme takes into account the dimension of argument construction as well as that of critique at the very first levels of the "learning progression" suggested by Osborne et al. (2016). Therefore, there has been little attention paid to refutation of ideas among students.

The transfer to practice analysis (Research Question 3) showed that the E-group was able to include argumentation as part of the activities in its teaching unit carried out during its

school practices to a greater extent and significantly more so than the C-group. Although the absolute values are not very high, it should be noted that PESTs are not forced to address argumentation during their school practices.

On a separate note, in this study the limitations and features of the tasks used for assessment (pre-test and post-test) were taken into account. These tasks focus on assessing PESTs' ability to provide full arguments, which is one of the aspects – of medium difficulty – highlighted in argumentation competence according to the "learning progression" approach (Osborne et al., 2016). Assessment tasks are designed with the aim that scientific knowledge involved in them should not pose an obstacle for PESTs to answer questions correctly. Nevertheless, this may have become a handicap to assess all of the potential progress that may have been made by the experimental group.

Finally, an interesting aspect to consider in future research is what Toulmin's model calls "backing". These are PESTs' previous ideas, judgments and beliefs on a certain socio-scientific or scientific problem, which have an impact on the arguments provided and the ability to assess or refute ideas. Another important line of work is to analyse how PESTs implement their knowledge on argumentation and their skills to analyse arguments in the design and development of activities and tasks for their students.

## Acknowledgments

This work is part of the "I+D Excelencia" project "Development and evaluation of scientific competences through context based and modelling teaching approaches" case studies (EDU2013-41952-P), funded by the Spanish Ministry of Economy and Finance through its 2013 research call.

## References

Andrews, R., & Mitchell, S. (2001). *Essays in Argument*. London: Middlesex Univ. Press.

Archila, P. A. (2012). La investigación en argumentación y sus implicaciones en la formación inicial de profesores de ciencias. [Argumentation research and its implications in science preservice teachers' training]. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, *9*(3), 361-375.

Belland, B. R., Glazewski, K. D., & Richardson, J. C. (2011). Problem-based learning and argumentation: Testing a scaffolding framework to support middle school students' creation of evidence-based arguments. *Instructional Science*, *39*(5), 667-694.

Berland, L. K., & Reiser, B. J. (2011). Classroom communities' adaptations of the practice of scientific argumentation. *Science Education*, *95*(2), 191-216.

Black, P., Harrison, C., Lee, C., Marshall, B., & William, D. (2003). *Assessment for Learning: Putting it into Practice*. Maidenhead, U. K.: Open University Press.

Boud, D., Cohen, R., & Sampson, J. (1999). Peer Learning and Assessment. *Assessment & Evaluation in Higher Education, 24*(4), 413-426.

Bulgren, J.A., Ellis, J.D. & Marquis, J.G. (2014). The Use and Effectiveness of an Argumentation and Evaluation Intervention in Science Classes. *Journal of Science Education and Technology, 23*(1), 82-97.

Bybee, R., & McCrae, B. (2011). Scientific Literacy and Student Attitudes: Perspectives from PISA 2006 Science. *International Journal of Science Education, 33*(1), 7-26.

Candela, A. (1999). *Ciencia en el aula. Los alumnos entre la argumentación y el consenso* [Classroom Science: Students between Argumentation and Consensus Construction]. Mexico: Paidós.

Cazden, C. (1991). *El discurso en el aula. El lenguaje de la enseñanza y el aprendizaje* [Classroom Discourse: The Language of Teaching and Learning]. Barcelona: Paidós-MEC.

Cebrián-de-la-Serna, M., & Monedero-Moya, J. J. (2014). Evolución en el diseño y funcionalidad de las rúbricas: Desde las rúbricas "cuadradas" a las erúbricas federadas [Evolution in the Design and Functionality of Rubrics: from Square Rubrics to Federated eRubrics]. *REDU Revista de Docencia Universitaria, 12*(1), 81-89.

Cebrián-Robles, D., Serrano-Angulo, J., & Cebrián-de-la-Serna, M. (2014). Federated eRubric Service to Facilitate Self-Regulated Learning in the European University Model. *European Educational Research Journal, 13*(5), 575-584.

Cebrián-Robles, D. (2016). CoRubric. http://corubric.com Accessed 27 December 2017.

Cetin, P. S. (2014). Explicit Argumentation Instruction to Facilitate Conceptual Understanding and Argumentation Skills. *Research in Science & Technological Education, 32*(1), 1-20.

Chao, K. L. & Jonassen, D. H. (2002). The effects of argumentation scaffolds on argumentation and problem solving. *Educational Technology Research and Development, 50*(3), 5-22.

Chin, C., & Osborne, J. (2010). Students' Questions and Discursive Interaction: Their Impact on Argumentation during Collaborative Group Discussions in Science. *Journal of Research in Science Teaching, 47*(7), 883-908.

Clark, D. B., & Sampson, V. (2008). Assessing Dialogic Argumentation in Online Environments to Relate Structure, Grounds and Conceptual Quality. *Journal of Research in Science Teaching, 45*(3), 293-321.

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

Custodio, E., & Sanmartí, N. (2005). Mejorar el aprendizaje en la clase de ciencias aprendiendo a escribir justificaciones [Improving Learning in Science Class by Learning to Write Justifications]. *Enseñanza de las Ciencias, extra number, VII Congreso*, 1-6.

Demircioğlu, T., & Uçar, S. (2012). The Effect of Argument-driven Inquiry on Pre-service Science Teachers' Attitudes and Argumentation Skills. *Procedia - Social and Behavioral Sciences, 46*, 5035-5039.

Deng, Y., & Wang, H. (2017). Research on Evaluating Chinese Students' Competence of Written Argumentation in the Context of Chemistry. *Chemistry Education Research*

*and Practice*, *18*(1), 127-150.

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

Duschl, R. (1998). La valoración de argumentaciones y explicaciones: promover estrategias de retroalimentación [Assessing Arguments and Explanations: Promoting Feedback Strategies]. *Enseñanza de las Ciencias*, *16*(1), 3-20.

Duschl, R.A., & Osborne, J. (2002). Supporting and Promoting Argumentation Discourse in Science Education. *Studies in Science Education*, *38*(1), 39-72.

Duschl, R., Ellenbogen, K., & Erduran, S. (1999). Understanding Dialogic Argumentation. *Annual meeting of the American Educational Research Association: On the* Threshold of the 21st Century: Challenges and Opportunities. Montreal (Canada): American Educational Research Association.

Erduran, S., & Jiménez-Aleixandre, M. P. (2008). *Argumentation in Science Education*. Berlin: 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.

Evagorou, M., & Osborne, J. (2013). Exploring Young Students' Collaborative Argumentation within a Socioscientific Issue. *Journal of Research in Science Teaching*, *50*(2), 209-237.

Felton, M., & Kuhn, D. (2001). The Development of Argumentative Discourse Skill. *Discourse Processes*, *32*(2-3), 135-153.

Franco-Mariscal, A.J. (2015). Competencias científicas en la enseñanza y el aprendizaje por investigación. Un estudio de caso sobre corrosión de metales en secundaria [Scientific Competences in Teaching and Learning through Research: a Case Study about the Corrosion of Metals in Secondary Education]. *Enseñanza de las Ciencias*, *33*(2), 231-252.

Furió, C., & Domínguez, C. (2007). Problemas históricos y dificultades conceptuales de los estudiantes en la conceptualización de sustancia y compuesto químico. [Historical problems and students' difficulties to the conceptualization of chemical substance and compound]. *Enseñanza de las Ciencias*, *25*(2), 241-258.

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

Henao, B. L., & Stipcich, M. S. (2008). Educación en ciencias y argumentación: la perspectiva de Toulmin como posible respuesta a las demandas y desafíos contemporáneos para la enseñanza de las ciencias experimentales [Education in Science and Argumentation: Toulmin's Perspective as a Potential Response to Modern Demands and Challenges in Experimental Science Teaching]. *Revista Electrónica de Enseñanza de las Ciencias*, *7*(1), 47-62.

Henderson, J. B., Osborne, J., MacPherson, A., & Szu, E. (2014). A New Learning Progression for Student Argumentation in Scientific Contexts. In C. P. Constantinou, N. Papadouris & A. Hadjigeorgiou (Eds.), *Proceedings of the ESERA 2013 Conference: Science Education Research for Evidence-Based Teaching and Coherence in Learning* (pp. 726-742). Nicosia (Cyprus): Springer.

Hennessey, G. (1991). *Analysis of Concept Change and Estatus Change in Sixth Graders'*

*Concepts of Force and Motion*. Doctoral Thesis. Madison: University of Wisconsin.

Hofstein, A., & Lunetta, V. N. (2004). The Laboratory in Science Education: Foundations for the Twenty-First Century. *Science Education, 88*(1), 28-54.

Iordanou, K., & Constantinou, C. P. (2014). Developing Pre-service Teachers' Evidence-based Argumentation Skills on Socio-scientific Issues. *Learning and Instruction, 34*, 42-57.

Jiménez-Aleixandre, M. P. (2002). Knowledge Producers or Knowledge Consumers? Argumentation and Decision Making About Environmental Management. *International Journal of Science Education, 24*(11), 1171-1190.

Jiménez-Aleixandre, M. P. (2005). Simposio la construcción del discurso científico socialmente contextualizado [Symposium on the Construction of the Socially-Contextualised Scientific Discourse]. *Enseñanza de las Ciencias, extra number, VII Congreso*, 1-6.

Jiménez-Aleixandre, M. P. (2010). *10 ideas clave. Competencias en argumentación y uso de pruebas* [10 Key Ideas: Argumentation Competencies and Use of Evidence] Barcelona (Spain): Graó.

Jiménez-Aleixandre, M. P., Bugallo, A., & Duschl, R. (2000). "Doing the Lesson" or "Doing Science" Argument in High School Genetics. *Science Education, 84*(6), 757-792.

Jiménez-Aleixandre, M. P., & Díaz, J. (2003). Discurso de aula y argumentación en la clase de ciencias: cuestiones teóricas y metodológicas [Classroom Discourse and Argumentation in Science Class: Theory and Method]. *Enseñanza de las Ciencias, 21*(3), 359-370.

Jiménez-Aleixandre, M. P., & Puig, B. (2010). Argumentación y evaluación de explicaciones causales en ciencias: el caso de la inteligencia [Argumentation and Assessment of Causal Explanations in Science: The Case of Intelligence]. *Alambique, 63*, 11-18.

Jonassen, H. D. (2004). *Learning to solve problems. An instructional design guide*. San Francisco, CA: Pfeiffer.

Justi, R. (2006). La enseñanza de ciencias basada en la elaboración de modelos [Teaching Science Based on Models]. *Enseñanza de las Ciencias, 24*(2), 173-184.

Kelly, G., & Takao, A. (2002). Epistemic Levels in Argument: an Analysis of University Oceanography Students' Use of Evidence in Writing. *Science Education, 86*(3), 314-342.

Kind, P. M., Kind, V., Hofstein, A., & Wilson, J. (2011). Peer Argumentation in the School Science Laboratory-Exploring Effects of Task Features. *International Journal of Science Education, 33*(18), 2527-2558.

Klein, P. D. (2004). Constructing scientific explanations through writing. *Instructional Science, 32*(3), 191-231.

Kovalainen, M., & Kumpulainen, K. (2005). The discursive practice of participation in an elementary classroom community. *Instructional Science, 33*(3), 213-250.

Kuhn, D. (1992). Thinking as Argument. *Harvard Educational Review, 62*(2), 155-178.

Kuhn, D. (1993). Science as Argument: Implications for Teaching and Learning Scientific Thinking. *Science Education, 77*(3), 319-337.

Kuhn, D. (2010). Teaching and Learning Science as Argument. *Science Education, 94*(5), 810-824.

Lemke, J. L. (1990). *Talking Science: Language, Learning, and Values*. Norwood, NJ:

Ablex.

Litman, C., Marple, S., Greenleaf, C., Charney-Sirott, I., Bolz, M. J., Richardson, L. K., & Goldman, S. R. (2017). Text-Based Argumentation with Multiple Sources: A Descriptive Study of Opportunity to Learn in Secondary English Language Arts, History, and Science. *Journal of the Learning Sciences, 26*(1), 79-130.

Llodra, J. C. (2012). Encuesta de Salud Oral en España 2010. [A Survey on Oral Health in Spain in 2010]. *RCOE, Revista del Ilustre Consejo General de Colegios de Odontólogos y Estomatólogos de España*, *17*(1), 13-46.

Martin, J. R., & Veel, R. (1998). *Reading Science*. Londres: Routledge.

Martínez, A., & Ibáñez, O. (2006). Resolver situaciones problemáticas en genética para modificar las actitudes relacionadas con la ciencia [Solving Issues in Genetics to Change Science-Related Attitudes]. *Enseñanza de las Ciencias*, *24*(2), 193-206.

Marttunen, M. (2002). Teaching Argumentation in Secondary School Through Computer Based and Face-to-face Debate. In M. Driscoll & T. Reeves (Eds.), *Proceedings of E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2002* (pp. 1862-1865). Chesapeake, VA: Association for the Advancement of Computing in Education (AACE).

Marttunen, M., & Laurinen, L. (2007). Collaborative learning through chat discussions and argument diagrams in secondary school. *Journal of Research on Technology in Education*, *40*(1), 109-126.

Marttunen, M., & Laurinen, L. (2001). Learning of argumentation skills in networked and face-to-face environments. *Instructional Science*, 29(2), 127-153.

McNeil, K. L., & Knight, A. M. (2013). Teachers' Pedagogical Content Knowledge of Scientific Argumentation: The Impact of Professional Development on K–12 Teachers. *Science Education*, *97*(6), 936-972.

Meyer, H., & Benavot, A. (2013). *PISA, Power, and Policy: The Emergence of Global Educational Governance*. Providence, RI: Symposium Books.

Ministerio de Educación [Spanish Ministry of Education] (2010). *Evaluación general de diagnóstico 2009 educación primaria. cuarto curso. informe de resultados* [General Assessment of Diagnosis 2009 Primary Education. 4th Year. Results Report]. Madrid (Spain): Ministerio de Educación.

Newton, P., Driver, R., & Osborne, J. (1999). The Place of Argumentation in the Pedagogy of School Science. *International Journal of Science Education*, *21*(5), 553-576.

Nielsen, J. A. (2012). Arguing from Nature: The Role of 'Nature' in Students' Argumentations on a Socio-scientific Issue. *International Journal of Science Education*, *34*(5), 723-744.

OECD (2006a). *PISA. Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006*. Annex A, Additional Science Units. Unit 2. Tooth Decay (pp. 127-130). Brussels: OECD.

OECD (2006b). *PISA. Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006*. Annex A, Additional Science Units. Unit 9. Lip Gloss (pp. 153-155). Brussels: OECD.

OECD (2016). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy*. Paris: OECD Publishing.

Oliva, J. M., Aragón, M. M., & Cuesta, J. (2015). The Competence of Modelling in Learning Chemical Change: A Study with Secondary School Students. *International Journal of*

*Science and Mathematics Education, 13*(4), 751-791.

Osana, H. P., & Seymour, J. R. (2004). Critical Thinking in Preservice Teachers: A Rubric for Evaluating Argumentation and Statistical Reasoning. *Educational Research and Evaluation, 10*(4-6), 473-498.

Osborne, J. F., Henderson, J. B., MacPherson, A., Szu, E., Wild, A., & Yao, S. (2016). The Development and Validation of a Learning Progression for Argumentation in Science. *Journal of Research in Science Teaching, 53*(6), 821-846.

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.

Özçinar, H. (2015). Scaffolding Computer-mediated Discussion to Enhance Moral Reasoning and Argumentation Quality in Pre-service Teachers. *Journal of Moral Education, 44*(2), 232-251.

Pinochet, J. (2015). El modelo argumentativo de Toulmin y la educación en ciencias: una revisión argumentada [Toulmin's argumentative model and science education: an argued review]. *Ciênc. Educ. Bauru, 21*(2), 307-327.

Ragonis, N., & Shilo, G. (2014). Drawing Analogies between Logic Programming and Natural Language Argumentation Texts to Scaffold Learners' Understanding. *Journal of Information Technology Education, 13,* 73-89.

Revel, A., Couló, A., Erduran, S., Furman, M., Iglesia, P., & Adúriz-Bravo, A. (2005). Estudios sobre la enseñanza de la argumentación científica escolar. [Studies on Teaching School Scientific Argumentation: Science Teaching]. *Enseñanza de las Ciencias, extra number, VII Congreso*, 1-5.

Rodríguez-Mora, F. (2016). *El "consumo de agua de bebida envasada" como contexto para el desarrollo de competencias científicas. Un estudio de caso en 3er curso de la educación secundaria obligatoria* ["Drinking Bottled Water" as a Context for Developing Scientific Competencies: a Case Study in 3rd Year of Compulsary Secondary Education]. Doctoral Thesis. Málaga (Spain): Universidad de Málaga.

Sadler, T., & D. Zeidler (2005). Patterns of Informal Reasoning in the Context of Socioscientific Decision Making. *Journal of Research Science Teaching, 42*(1), 112-138.

Sampson, V., & Blanchard, M. R. (2012). Science Teachers and Scientific Argumentation: Trends in Views and Practice. *Journal of Research in Science Teaching, 49*(9), 1122-1148.

Sanmartí, N. (coord.) (2003). *Aprendre ciències tot aprenent a escriure ciència*. Barcelona (Spain): Edicions 62.

Sheiham, A. (2005). Identification of appropriate outcomes for an ageing population. *Bulletin of World Health Organization, 83*(9), 644-645.

Simonneaux, L. (2001). Role-Play or Debate to Promote Students' Argumentation and Justification on an Issue in Animal Transgenesis. *International Journal of Science Education, 23*(9), 903-927.

Starling, T., & Lee, H. (2015). Synchronous online discourse in a technology methods course for middle and secondary prospective mathematics teachers. *Contemporary Issues in Technology and Teacher Education, 15*(2), 106-125.

Sutton, C. (1992). Words, Science and Learning. Buckingham: Open University Press.

Sutton, C. (1997). Ideas sobre la ciencia e ideas sobre el lenguaje [Ideas about Science and Ideas about Language]. *Alambique, 12,* 8-32.

Taylor, C. A. (1996). *Defining Science: A Rhetoric of Demarcation*. Wisconsin: The University of Wisconsin Press.

Toulmin, S. E. (1958). *The Uses of Argument*. Cambridge: Cambridge University Press.

Toulmin, S. E. (1972). *Human Understanding: Vol. 1. The Collective Use and Evolution of Concepts*. Princeton. NJ: Princeton University Press.

Toulmin, S. E. (2003). *Return to Reason*. Cambridge: Harvard University Press.

Toulmin, S. E., Rieke, T., & Janik, A. (1979). *An Introduction to Reasoning*, New York: Macmillan.

Tsai, C. (2015). Improving Students' PISA Scientific Competencies through Online Argumentation. *International Journal of Science Education*, *37*(2), 321-339.

Tüysüz, C., Demirel, O. E., & Yildirim, B. (2013). Investigating the Effects of Argumentation, Problem and Laboratory Based Instruction Approaches on Pre-Service Teachers' Achievement Concerning the Concept of "Acid and Base". *Procedia - Social and Behavioral Sciences*, *93*, 1376-1381.

Visser, J., Barach, P., Van Breda, J., & Visser, Y. L. (2007). *Building the scientific mind. Learning in the perspective of complex and long-term change*. Eyragues (France): Learning Development Institute.

Walton, D. N. (1989). Dialogue Theory for Critical Thinking. *Argumentation*, *3*(2), 169-184.

Walton, D. N. (1990). What is Reasoning? What is an Argument?. *The Journal of Philosophy*, *87*(8), 399-419.

Yaman, F. (2017). Effects of the Science Writing Heuristic Approach on the Quality of Prospective Science Teachers' Argumentative Writing and Their Understanding of Scientific Argumentation. *International Journal of Science and Mathematics Education*, 1-22, doi: 10.1007/s10763-016-9788-9

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.

## Annex I

### Question 1: Tree Growth

*"In the course of their development, woody stem plants grow in height and width for years. The growth in width marks the tree trunk with a ring for each year. If the trunk is cut transversely, as shown in the figure below, each of these rings can be seen with a different thickness, depending on the weather conditions of the year in question (Figure 4). Look carefully at the tree growth rings in the figure and answer the following question: What ring corresponds to the year when weather conditions were most favourable for the tree? Justify your answer by providing evidence".*

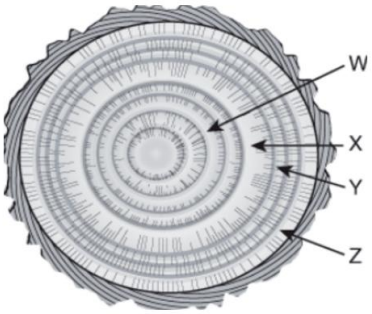


**Fig. 4** Image in the question on tree growth (Proposal of the general diagnosis assessment 2009 for Primary Education from the Spanish Ministry of Education, 2010)

**Question 2: Purity of Bottled Water**

*"This is a label found on a bottle of a bottled water brand (front and back). Read it carefully: (Figure 5). Do you think this bottled water is "pure, truly pure water" as the label reads? Justify your answer based on the information (terms or data) given on the label".*

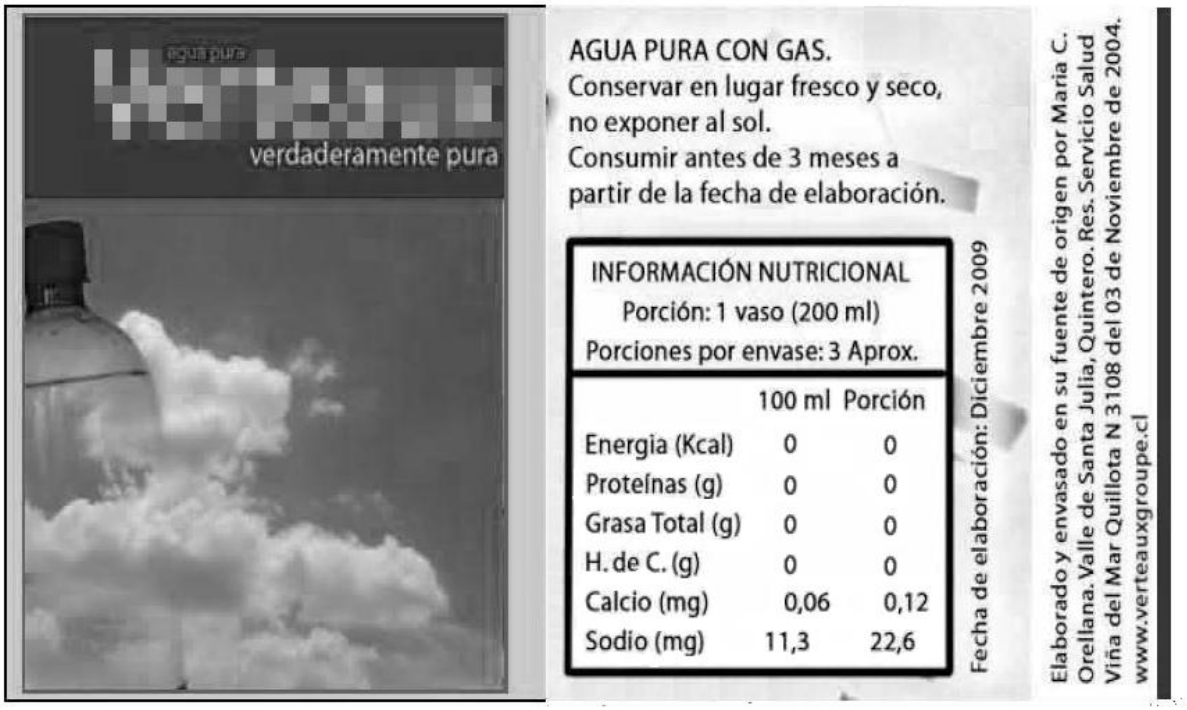


**Fig. 5** Image of the purity of bottled water activity (Proposal in Rodríguez-Mora's Doctoral Thesis, 2016)

**Annex II**

**Question 3: Tooth Decay**

*"Do bacteria play an important role in tooth decay? Justify your answer".*

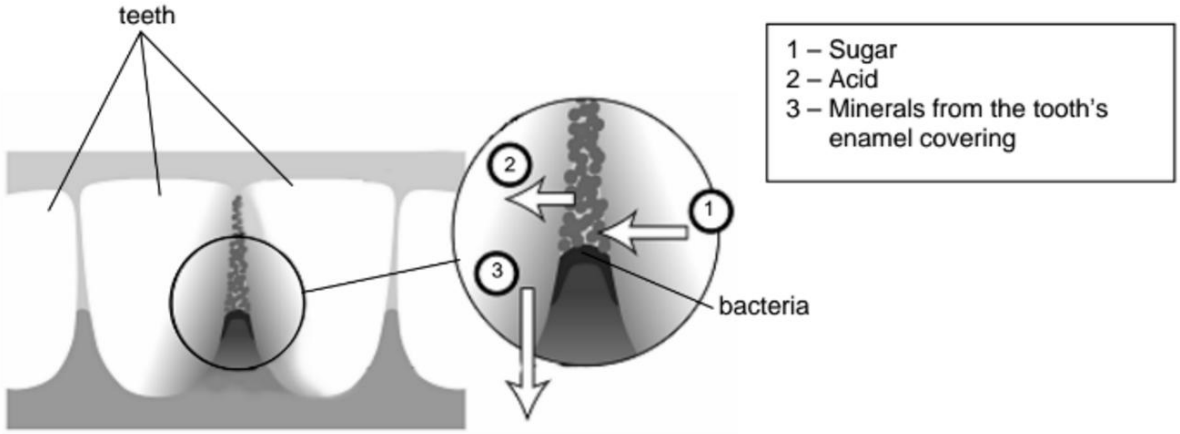


**Fig. 6** Image in the tooth decay activity (OECD, 2006a)

**Annex III**

**Question 4: Lipstick Hardness**

*"Can we change the recipe to make it softer? Justify your answer by providing evidence"* (table 4).

**Table 4** Information given in the activity on lipstick hardness (OECD, 2006b)

| Lip gloss | Lipstick |
|---|---|
| **Ingredients:** | **Ingredients:** |
| 5 g castor oil | 5 g castor oil |
| 0.2 g beeswax | 1 g beeswax |
| 0.2 g palm wax | 1 g palm wax |
| 1 teaspoon of colouring substance | 1 teaspoon of colouring substance |
| 1 drop of food flavouring | 1 drop of food flavouring |
| **Instructions:** | **Instructions:** |
| Heat the oil and the waxes in a container placed in hot water until you have an even mixture. Then add the colouring substance and the flavouring, and mix them in. | Heat the oil and the waxes in a container placed in hot water until you have an even mixture. Then add the colouring substance and the flavouring, and mix them in. |