

The final publication is available at link.springer.com

(<https://link.springer.com/article/10.1007/s11165-020-09985-z>)

An Assessment Using Rubrics and the Rasch Model of 14/15-Year-Old Students' Difficulties in Arguing About Bottled Water Consumption

**Francisco Rodríguez-Mora; Daniel Cebrián-Robles & Ángel
Blanco-López**

Francisco Rodríguez-Mora franrodriguez@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

An assessment using rubrics and the Rasch model of 14/15-year-old students' difficulties in arguing about bottled water consumption

Abstract: The development of students' argumentation competence is one of the major aims of science education and everyday problems, such as decisions on the consumption of tap water or bottled water, may be suitable contexts to encourage it. Citizens consider of interest certain controversies associated with the bottled water consumption as an alternative to tap water, which has a significant effect on several aspects of their daily lives. Deciding whether or not to drink bottled water can be used as an interesting context for science teaching activities. To make appropriate decisions on this issue, students must know how to argue with scientific rigor. This paper presents an assessment of the degree of performance of 14/15-year-old students in scientific argumentation within this context. A specific rubric has been designed for each of the assessment activities used, and the Rasch model has been applied to analyse the degree of difficulty that each of the essential elements of an argument (evidence, justification, and conclusion) has for the students. Thus, evidence is generally shown to be the most difficult element of the argument for students followed by justifications and finally conclusions. The analysis procedure used can help to establish a sequence of argumentation activities within a particular context. Finally, the results obtained can be used in teaching to prepare the appropriate assistance according to the argumentation activity at hand.

Keywords: argumentation; water consumption; student performance; Rasch model

Introduction

As the literature has shown, helping students develop their argumentation competence is a complex issue (Bravo-Torija & Jiménez-Aleixandre, 2018; Osborne et al., 2016). It is complicated for different reasons, among them, the difficulty of the specific activities that are posed to students, many of which are due to the scientific knowledge involved (Berland & McNeill, 2010), the nature and number of evidences needed to argue (Bravo-Torija & Jiménez-Aleixandre, 2018) and finally, the context of the argument

itself (Simonneaux, 2008). Three fundamental aspects should be considered when assessing argumentative performance: the instructional framework, the argumentative product and the argumentation process (Berland & McNeill, 2010), which are assessed and considered in progressive learning proposals in different ways by different authors. Other authors analyse in the argumentative products the use of evidence for decision making in different contexts (Bravo-Torija & Jiménez-Aleixandre, 2018). The importance of connecting science through carefully selected concepts and principles for students to see their application in their daily lives is widely recognised nowadays (King & Ritchie, 2012). It is important to improve the capacity to detect the bias that can occur in certain arguments and, therefore, to avoid being too partisan to achieve a balanced perspective in the argument (Walton, 1989). Today, citizens are educated to make critical readings of the media and scientific news appearing in the everyday context (Oliveras, Márquez, & Sanmartí, 2013), such as those related to the controversy between the consumption of tap water or bottled water, for instance (Doria, 2006). Rodríguez-Mora (2016) substantiates the importance of citizens' decision on whether or not to drink bottled water and how its treatment in the teaching of sciences can be used to help students consider this problem. This demands students to argue for or against its consumption.

In this sense, this paper presents a procedure to assess the difficulty of specific argumentation activities on the consumption of bottled water, providing a more objective performance assessment of Compulsory Secondary Education students which makes it possible to identify the specific elements of the arguments that are most difficult for students and to provide appropriate teaching aids when necessary (Erduran, 2007). This procedure combines specific assessment rubrics (Cebrián-Robles, Franco-Mariscal, & Blanco-López, 2018) and the Rasch model (Masters, 1982) that has

already been successfully tested in other science teaching works (Hsieh, Lin, Liu & Tsai, 2019)

The following research questions are posed:

- A. Which elements (evidence, justification and conclusion) are the most difficult for students in argumentation activities about the consumption of bottled water?
- B. To what extent does the explicit inclusion of evidence in the description of an activity influence the difficulty of argumentation?
- C. To what extent does the scientific knowledge involved in an activity influence the difficulty of argumentation?

Literature review

Scientific argumentation

There are different ways of understanding and defining scientific argumentation. Osborne et al. (2016) consider that the argumentation “is not simply an aptitude that can be assessed, but rather, a *competency* which draws on a mix of content knowledge, procedural knowledge, and epistemic knowledge (OECD, 2012). The construction of an argument requires the ability to remember appropriate information and to construct a justified relationship between this and the claim.” (p. 3). Jiménez-Aleixandre & Puig-Mauriz (2010) consider that for there to be argumentation there has to be knowledge subject to assessment, in our case scientific knowledge, and evidence to confirm or refute it. The review of the literature on argumentation by Pinochet (2015) reflects the importance of learning argumentation skills, since it improves science education processes and their culture. Besides, this review also reveals difficulties and challenges for teaching scientific argumentation. One of the primary difficulties and

challenges seen in scientific argumentation is the lack of the scientific knowledge necessary to argue based on evidence (Henderson, McNeill, González-Howard, Close, & Evans, 2018; Grooms, Sampson & Enderle, 2018). On the one hand, the absence or lack of scientific *backing* to understand the depth of a problem may drive students to justify themselves based on guesswork and, failing to incorporate multiple justifications or explanations can result in their omission of the implied backing (Bell & Linn, 2000). This situation can further lead to the creation of fallacies due to false beliefs rooted in the student body, such as problems with the validity of claims; naïve conceptions about the argumentative structure; the effects of beliefs over claims based on evidence; inadequate sampling of evidence; and alteration of the argument and evidence leading to biased results (Zeidler, 1997). The importance of beliefs in argumentation has been highlighted (Zeidler, 1997; Godden, 2010). So, Bravo-Torija & Jiménez-Aleixandre (2018) said that “Evagorou, Jiménez-Aleixandre and Osborne (2012) examined students’ difficulty in identifying counterevidence to their existing beliefs, even in cases when their choices contradicted the available evidence” (p. 623). Furthermore, the use of evidence in arguments presents many difficulties for students. Bravo-Torija & Jiménez-Aleixandre (2018) analysed these difficulties, among which are: the interpretation of data and the establishment of connections between the different datasets; the inclusion of the evidence within justifications; and the demonstration of an appropriate level of understanding to interpret the evidence and thus assess different options from the point of view of the strengths and weaknesses of each one. Students also tend to make explanations without evidence or use single pieces of evidence in the argument rather than a set of them (Evagorou & Avraamidou, 2008). Teachers are, therefore, required to strive to design learning activities that encourage the use of evidence in arguments within training programmes (Bravo-Torija &

Jiménez-Aleixandre, 2018). Henderson et al. (2018) addressed these difficulties by proposing a series of challenges in their work, including: establishing and fostering an argumentative culture in class; assessing arguments in a valid and reliable manner; encouraging and supporting the collaborative and social discourse of students; raising awareness about the professional development of teachers in argumentation; and ascertaining whether the failure to argue is due to lack of knowledge or whether, epistemologically, they do not know how to argue.

Another important aspect is the contexts in which argumentation can be considered, which can be theoretical, empirical, decision making, confirmation of predictions or critical evaluation of statements. In this sense, teaching activities and strategies should be designed and proposed in appropriate contexts that enable students to understand and use scientific discourse and models. They should also be able to handle science-related situations from a critical perspective that allows them to provide a solution and establish a debate on authentic problems (Jiménez-Aleixandre et al., 2000), that is, interesting and significant problems for students. The so-called socio-scientific issues stand out among these issues (Sadler & Zeidler, 2005).

The Toulmin model (1958) has long been used as a theoretical framework to analyse the argumentation product in science class (Jiménez-Aleixandre & Díaz-de-Bustamante, 2003), as it offers a valuable approach for analysing the elements of students' arguments when attempting to substantiate their conclusions (Pinochet, 2015). However, some authors have argued that this Toulmin model has limitations, especially for the oral or dialectical dimension of argumentation (Nielsen, 2012).

This model gathers the generic structure that any argumentative text should have, and identifies its fundamental components, as well as the order in which they should appear

in discourse. To this end, the scheme proposed by Toulmin (1958) has six fundamental components to assess the quality of argumentation, each of which has a certain function in the argumentative discourse as follows: claim, data, warrant, backing, qualifiers, and rebuttals. These elements are known as the Toulmin's Argument Pattern (TAP) and according to Erduran (2007) in addition to a claim, the argument needs: "data that support that claim; warrants that provide a link between the data and the claim; backings that strengthen the warrants; and finally, rebuttals which point to the circumstances under which the claim would not hold true" (p. 57). The TAP has been used by Capkinoglu, Yilmaz & Leblebicioglu (2019) to evaluate the quality of argumentations presented by students about to local socio-scientific issues. However, some works such as those of Osborne et al. (2016) focus on claim, evidence and warrant, with the understanding that: "Taken as a whole, a competency for scientific argumentation demands a complex orchestration of construction and critique of claims, warrants, and evidence in situations that require scientific knowledge to resolve" (Osborne et al., 2016, p. 6). Cebrián-Robles, Franco-Mariscal & Blanco-López (2018) have named these three elements of the Toulmin's model as 'justification' (instead of warrant in Toulmin's model), 'evidence' (instead of data) and 'conclusion' (instead of claim).

The use of the TAP model has allowed different authors, such as Capkinoglu, Yilmaz & Leblebicioglu (2019), or Erduran (2007) among others, to analyze the quality of the arguments of different groups of students. This establishes five levels in which the elements of the Toulmin (1958) model are combined. However, other works deal more analytically with the assessment of the arguments by differentiating the components of such a scheme (Belland, Glazewski & Richardson, 2011; Osborne et al. 2016) understanding that the difficulties such elements present for students may be very

different. For example, Bravo-Torija & Jiménez-Aleixandre (2018) have found that students' difficulty in using evidence in their arguments has been attributed to a lack of scientific knowledge (Berland & McNeill, 2010).

Use of rubrics to assess the arguments

Rubrics are well-established assessment instruments in the educational literature for grading the level of achievement (Stevens & Levi, 2005) and also for researching new learning progression of argumentation competence (Osborne et al., 2016). In the assessment contexts, its use allows students to clarify what is important and helps them to know the strengths and weaknesses of their arguments (Lu & Zhang, 2013). Other authors such as Jonsson & Svingby (2007) point out that the rubrics have benefits for the assessment: “increased consistency of scoring, the possibility to facilitate valid judgment of complex competencies, and promotion of learning” (p. 131). Therefore, this can be very useful to assess the students' ability to argue (Osborne et al. 2016) in a more analytical way, taking into account each of the components of the arguments.

In secondary education, rubrics have been used to assess arguments, as in the work of Belland, Glazewski & Richardson (2011) in which numerical scores for the claim, evidence, and connection of claim to evidence quality were assigned, or in the work of Deng & Wang (2017) in which, in addition to these elements, the rebuttal was also evaluated.

The Rasch model in Science Education

The Rasch model (1960) is a quantitative analysis based on the item response theory models which is part of a group of mathematical test validation models. It is considered that this model allows a good adjustment of the data and the joint measurement of people and items in the same dimension or construct. The Rasch model according to

Osborne et al. (2016): “provide convenient and rich ways to model both person proficiency and item difficulty on the same scale” (p. 10) and has been used by these authors to propose Learning Progression for Argumentation in Science based on the elements of an argument among other aspects.

The Rasch model has been used in conjunction with the rubrics in the area of argumentation as a piece of validity evidence for learning progression-based assessment items (Gotwals & Songer, 2013). Lee et al. (2014) use them for the assessment of progressive learning of scientific argumentation, but focused on uncertainty-infused scientific argumentation and, at the same time, He et al. (2020) use them to improve the quality of questionnaire items with argumentation questions for high school students.

Instructional context

This research was carried out in the context of a teaching proposal entitled “*Is it necessary to drink bottled water?*”, aimed at 25 students of 9th grade (14-15 years old) studying Physics and Chemistry. The sequence required 13 one-hour sessions for implementation in the classroom.

This proposal seeks to bring specific controversies associated with the consumption of bottled drinking water as an alternative to public mains water (tap water) (Ferrier, 2001; Gleick & Cooley, 2009) to the sphere of science education. And for that, it combines the processing of knowledge of chemistry with decision making (Siribunnam, Nuangchalerm, & Jansawang, 2014). To this end, students were asked two key questions that constitute the backbone of the classroom activities: *Do we have to consume bottled water?* and *Is bottled water better than tap water?* These two key questions were broken down into other more specific ones with the aim of organising the development of the sequence of activities in the classroom (see Table 1).

[INSERT HERE TABLE 1]

The teaching proposal aimed to develop students' scientific competences following the scheme proposed by PISA (OECD, 2006): identification of issues, elaboration of explanations, and use of scientific evidence. In particular, according to OECD (2006) “using scientific evidence includes accessing scientific information and producing arguments and conclusions based on scientific evidence” (p. 30). In the teaching proposal the argumentation was focused on the three aspects under consideration: the recognition of drinking water as an example of dissolution, the differences in chemical composition between tap and bottled water, and the reasons for deciding which water is the best. From the Toulmin model mentioned above, these argumentation activities required the student to identify evidence, develop justifications and draw conclusions. Once the teaching proposal was finished, the students completed an assessment questionnaire that included the three argumentation activities from which the data used in this research have been taken.

Methods

Data collection instruments and analysis tools

Once the teaching proposal was completed, three argumentation activities were used for data collection in this research. And for each of them, a rubric as instruments to analyse the degree of student performance was designed. These rubrics, specific to each of the assessment activities, are the result of adapting a basic argumentation rubric to the characteristics of each activity being evaluated (Cebrián-Robles, Franco-Mariscal, & Blanco-López, 2018).

Activity 1. Is bottled water pure?

This activity is related to the characterisation of bottled water as a solution. The belief in the “purity” of bottled drinking water –as opposed to tap water— is one the water’s main identifying marks for consumers, and one of the main reasons for their consumption (Ferrier, 2001). The presence of the phrase *Pure Water, truly pure* in a bottled water label and the difference in meanings of this term in colloquial or scientific language, are used in this activity. Students are expected to interpret the data provided by the chemical analysis seen on the label to justify this phrase from a chemical point of view. Its development, therefore, requires basic chemical knowledge of homogeneous mixtures of substances (solutions) and the chemical concept of pure substance (Figure 1).

[INSERT HERE FIGURE 1]

The specific rubric to assess the quality of the arguments given by the students is shown in Figure 2.

[INSERT HERE FIGURE 2]

An appropriate response to this question, in order to achieve a maximum level in all sections of the rubric of this assessment activity would be something in the lines of: *“The bottled water on the label is not chemically pure, because it contains several dissolved substances. It contains gas, calcium and sodium that make this bottled water a homogeneous mixture.”* This answer includes two pieces of evidence provided by the label *“Sparkling water”* and *“Contains calcium and sodium”*; a justification *“Not chemically pure because it contains several dissolved substances. Bottled water is a homogeneous mixture”* and a conclusion *“This bottled water is not chemically pure”*.

The letter "S", followed by a number corresponding to the alphabetical order, is used to

identify the students, while the assessment of their answers is indicated by a code corresponding to the levels of the rubric, which are represented by a number followed by a letter (E -evidence-, J -justification- or C –conclusion-), for each of the elements of the argument. Thus, an argument assessed as 4E5J5C means the highest level of achievement.

Activity 2. Does bottled water have lime?

This activity is related to the differences in chemical composition between tap water and bottled water. It refers to one of the most widespread beliefs among students: bottled water does not contain “lime” (it is not hard); so, the presence of lime in tap water appears as one of the most important reasons to justify the consumption of bottled drinking water instead of tap water, according to different preliminary studies developed in this research (Rodríguez-Mora & Blanco-López, 2012).

The hardness of tap water –a term related to the concentration of calcium and magnesium ions in chemistry– is one of the most common basic parameters used to determine the quality of drinking water, as it largely determines its taste, among other features. Although the consumption of hard water does not generally pose a problem for people’s health (WHO, 2011), it is usually perceived as a negative aspect linked to low quality and potential health risks and, therefore, viewed as a determining factor in the choice of bottled water as an alternative to public tap water (Doria, 2010).

The use of hard water in daily life generates whitish incrustations commonly known as “lime”. Therefore, the lime found in water is a well-known term, frequently used in student’s daily lives to refer to the hardness of water, which is why the colloquial term was used in the teaching proposal and during the activities.

The activity was presented in the following manner: “Argue the veracity or falsity of the following statement: ‘Bottled water has no lime. Only tap water has lime.’ State as many reasons as you can to complete your answer.” Students were expected to evaluate the statement in the light of scientific knowledge and provide adequate evidence as to the veracity or otherwise of the proposed conclusion. This activity required students to know the chemical fundamentals of water hardness and its relationship with the presence of calcium and magnesium ions in a solution, fundamentally. The rubric shown in Figure 3 was proposed to assess students’ responses:

[INSERT HERE FIGURE 3]

The answer considered adequate of this assessment activity (Figure 3) would be in the lines of “*Bottled water also has lime, as it contains calcium and magnesium in its composition. These salts are primarily responsible for the hardness of water, which we colloquially refer to as “lime” in water.* Students are expected to provide proof that “Bottled water has calcium and magnesium”; a justification such as “*The calcium and magnesium salts we find in water are primarily responsible for water hardness, which we colloquially refer to as 'lime' in water*” and the conclusion similar to “It is false, *bottled water also contains lime*”.

Activity 3. Is high mountain water the same as tap water?

This activity is related to the characterisation of the reasons on which is the best type of water: tap or bottled. A person's body water supply is renewed continuously at a rate that depends on the amount of water ingested, but it can be said that 99% of the body's water supply is renewed every 50 days (Péronnet et al., 2012). The bottled water industry has used this physiological fact of continuous renewal of water in our body as the cornerstone of some advertising campaigns, and certain brands have used “this

perpetual renewing flow of water in our bodies” to invite us to drink bottled water, water carrying “wonderful” properties thanks to its composition and origin (Gavidia-Catalan, 2009).

According to the above, a real advertisement from a well-known brand of bottled drinking water, published in a national newspaper was chosen to present this third activity to students. The images inform the students of the need to drink water due to a continuous process of “renewal” of this liquid in our organism: *“Up to 60% of your body is water that is renewed every 6 weeks. Would you renew yourself with just any water (...)? Renew the water in your body with L, pure high mountain water”*. Students were asked to answer the following question in a justified way (Figure 4).

[INSERT HERE FIGURE 4]

Based on the idea that we need to drink water for our body to develop its functions correctly, students were expected to propose different reasons (evidence) that justify or refute the validity of the advertising message. The assessment rubric used in this activity is shown in Figure 5.

[INSERT HERE FIGURE 5]

A response considered adequate is “I would renew myself equally by drinking tap water or high mountain water since both contain the minerals necessary to renew the water inside me”. Students were expected to provide the following evidence: “High mountain water contains minerals”, “Tap water contains minerals” and/or “We need water with minerals to renew the water inside ourselves”; a justification such as “Since both types of water have minerals and I need these to renew myself, they would be equally valid” and the conclusion in the lines of “I would renew myself equally by drinking high

mountain water or tap water”.

Using these rubrics, the students’ responses were analysed independently by two of the authors participating in this project. The reliability of the categorisation (allocation of responses to different levels of the rubrics) was examined by comparing the assessments of the authors of the work for the same answers. The research team agreed by consensus on the categorization of responses where there was no agreement (Escobar-Pérez & Cuervo-Martínez, 2008).

An example of student response and how it was assigned to the different levels of the rubric is shown below. The following answer was given by student S05 to the activity 1 (Figure 6):

[INSERT HERE FIGURE 6]

This student’s response is considered as inadequate, because it does not provide any evidence in his argument, so the level assigned is 1E. In addition, his justification is considered level 2J, when he says that “*pure water in chemistry is formed only by one substance*”, but there is no link between the evidence and the conclusion. However, he makes an accurate but not entirely adequate conclusion, which is assigned level 4C, because it the statement “*I think that, in scientific terms, it is not pure*” is not entirely explicit.

Hypotheses on the level of difficulty of these activities

In a previous study (Cebrián-Robles & Rodríguez-Mora, 2018), a hypothesis on the difficulty of these activities was considered based on the evidence needed to make an adequate scientific argument and the scientific knowledge involved in each of them. Taking into account these two aspects, the following hypotheses, related to research questions 2 and 3 respectively, were raised (Table 2):

a) the difficulty of the argumentation activity increases if the evidence does not appear in the description of a given activity (research question 2).

b) the difficulty increases if the necessary scientific knowledge has not been taught to the students during the teaching proposal (research question 3).

[INSERT HERE TABLE 2]

Since these assessment activities were completed after the implementation of the teaching proposal described above (Rodríguez-Mora, 2016), the chemical knowledge required to respond to each of them was considered as an essential variable when analysing their degree of difficulty. Thus, the activities should have a lower level of difficulty when the scientific knowledge involved in them had been explicitly taught during the teaching proposal.

Three levels of difficulty were hypothesised combining both criteria (availability of evidence and chemical knowledge), so the evaluation activity *Is bottled water pure?* would be the simplest (level 1) and the *Is high mountain water the same as tap water?* the most complex (level 3), as shown in Table 2.

Statistical processing

Once the students' answers were categorised according to the levels of achievement of each rubric, the degree of performance in the argumentation activities and the complexity differences of these activities. The levels of achievement of each rubric are analysed independently (0 if the student fails and 1 if he/she succeeds at least in this level of achievement) in the Rasch model and they were calculated using the one-dimensional Rasch model for answers scored in two or more ordered categories (Masters, 1982), which allowed analysing the questions on the same scale. This

one-dimensional model has been selected because Barnes & Wise (1991) have shown that, in addition to small samples, it remains robust. In this case, the student reliability must be greater than 0.5 in good agreement with the Rasch model (Boone et al., 2013, p. 230).

In this research, this model allowed converting the individual scores assigned to the students in the specific assessment rubrics to a logical scale, placing the position of the difficulty of the activities and the estimation of the individual ability of the students on a single scale using equal intervals and relative differences between the scores obtained. It was thus possible to visualise where the activities were placed in terms of difficulty with respect to the students in a relative and escalated manner (Masters, 1982). The R¹ calculation program was needed for this analysis. This program allows, among other things, presenting the results, through the TAM, eRm and WrightMap libraries, in a WrightMap graph, designed for simple psychometric studies on categorical data, according to its order of difficulty, showing the activities ordered by their difficulty on the right and the normal curve of students' responses on the left (Wright & Stone, 1979). This permitted assessing the difficulty of such activities for the selected sample of students.

Results

Student responses (n= 75) were analysed applying the specific rubrics for each activity (Figures 2, 3 and 5). As can be seen in Table 3, the distribution of achievement levels attained in *Evidence* is very similar in activities 1 and 3, while in activity 2 most of the students were at level 1. Something similar happens concerning *Justification*, students reached almost all the levels of achievement simultaneously, except for activity 2, in

¹ The R Project for Statistical Computing <https://www.r-project.org/>

which they did not exceed level 3. Students showed that both tap and bottled water contain mineral salts, a knowledge that had been addressed in the teaching proposal. However, the idea of water hardness was not reflected in any of the justifications, so none reached level 4. The conclusion was one of the elements of the argument that scored the highest, with a high percentage of responses at the highest levels of achievement (3, 4 and 5) for the three activities.

[INSERT HERE TABLE 3]

The most frequent applications of the Rasch model use dichotomous values (Osborne et al., 2016). Thus, the criteria used to convert the rubric scores into dichotomous values (adequate 1 and not adequate 0) were decided, considering the results shown in Table 3. Thus, concerning *Evidence*, appropriate responses have been found to be those that attain at least a level 3 in the rubrics (Figures 2, 3 and 5), since it is understood that responses at this level or higher provide relevant evidence even though they are not sufficient to answer the question posed. The same has been applied for *Justification* purposes, considering adequate justification from level 3 upwards. This level implies that the justification provided links evidence to the conclusion in a relevant manner even if it does not do so with scientific rigour. For the *Conclusion*, which turned out to be the simplest element of the arguments (Table 3), the adequate responses of the students should be those that reach level 4 at least. This would indicate that the correct response has been provided with scientific rigour, albeit not in an exhaustive manner.

In this way, the results shown in Table 4 are obtained as the product of the transformation of the frequencies for each of the levels of the rubrics (Table 3) into dichotomous values (adequate or not adequate) to be analysed with the Rasch model.

[INSERT HERE TABLE 4]

These data were used in the Rasch model to construct the *WrightMap* graph shown in Figure 7.

[INSERT HERE FIGURE 7]

In Figure 7, the first column represents the curve showing the distribution of the performance of the students in carrying out the three assessment activities. This represents the frequency of students in the horizontal axis, being higher its values the more to the left and the performance in the vertical axis, being more difficult the more upwards. In the successive columns are represents the results obtained for each of the three elements of the arguments. The coding used corresponds to two characters *XY*, where *X* represents the assessment activity (Activity 1. *Is bottled water pure?* Activity 2. *Does bottled water have lime?* or Activity 3. *Is high mountain water the same as tap water?*), while the character *Y* represents the essential element of an argument (*E* -evidence-, *J* -justification- or *C* -conclusion-). The reliability of the student responses for the three activities was 0.543, and since this value is greater than 0.5 (Boone et al., 2013, p. 230), which makes the data valuable for analysis.

It shows that the evidence is the most challenging element of an argument for the students, because in general they are located higher than the other elements (figure 7) and this, according to the Rasch model, represents a greater difficulty for the students. However, in the specific case of activity 3, the conclusion (3C) presents a greater difficulty than the evidence (3E). A lower difficulty level shows the justifications since two of them are in the central part (1J and 3J) and one on a lower level (2J) of Wright's map. Furthermore, it is clear that most students can give a correct conclusion, except in the case of activity 3 *Is high mountain water the same as tap water?*

Conclusions and teaching implications

With regard to the first research question, the results obtained show that the provision of evidence is the element of the argument that represents the greatest difficulty for students. This is in accordance with what is stated in the literature, in general, for all types of scientific argumentation contexts (Bravo-Torija & Jiménez-Aleixandre; 2018). The justifications are at an intermediate level of difficulty and the conclusions at the lowest level, except in the case of activity 3 *Is high mountain water the same as tap water?*, which may be due to the complexity of the activity for students of these ages, in particular maybe to the difficulty of the concept of water renewal in the human body.

With regard to the second research question, we consider that the hypothesis a) (Table 2) is accepted, since the results show the influence of the presence or not of evidence in the description of the activity on the difficulty of the argumentation task involved in it, so, this hypothesis anticipates the best performance grades obtained in the *Is bottled water pure?* activity compared to *Does bottled water have lime?* These results reaffirm those found in the literature regarding students' difficulty to provide evidence (Bravo-Torija & Jiménez-Aleixandre; 2018). By including evidence in the description of the activity, the quality of the arguments can be improved, to make up for the absence or lack of scientific knowledge that might lead to the formulation of justifications based on conjectures to give explanations, as Bell and Linn (2000) explain in their work.

The same does not apply to the chemical knowledge (research question 3, hypothesis b), Table 2). Thus, in the activity *Does bottled water have lime?*, in which the scientific knowledge necessary to answer it was taught during the teaching proposal, it has been found, contrary to expectations, to have lower performance than in the activity *Is high mountain water the same as tap water?* in which scientific knowledge had not been

taught (Table 4), This may be due to the complexity of the *Does bottled water have lime?* assessment activity primarily focused on the students' difficulty to provide evidence (Table 4). These results, which are different from those obtained by Grooms, Sampson & Enderle (2018) with students enrolled in a high school chemistry course (age 15–17) can be explained, as Zeidler (1997) indicated, as one of the typical errors of student argumentation, in which the effects of students' beliefs are more convincing than those that are contrary to their beliefs. This situation leads to weak arguments that compromise their ability to construct and assess their arguments. The teaching proposal, although explicitly addressing the concept of water hardness, was unable to help students overcome their initial –albeit deeply-rooted– beliefs about the presence of lime in water (Rodríguez-Mora & Blanco-López, 2012).

The analysis methodology used in this research allows testing the initial hypothesis and sequencing the activities to increase the performance of the students in all of them. The Rasch model allows working with small samples, so it is suitable for research like the one presented here. Combining the specific assessment rubrics and the Rasch model can help establish a sequence of argumentation activities within a particular context, considering the degree of difficulty for students both from a general point of view and each of the elements of the arguments according to the simplified Toulmin scheme used in this research.

The results and conclusions obtained should be considered with caution due to the small number of students who received the teaching proposal. Likewise, it should be noted that two aspects (evidence and scientific knowledge) have been explored that influence students' capacity to argue, while other aspects of the scientific argumentation, such as their understanding of this scientific practice itself (Grooms, Sampson & Enderle,

2018), are also important.

In accordance with the objectives of this research, the results obtained can be used in teaching to prepare the appropriate assistance according to the argumentation activity at hand. For instance, in the *Does bottled water have lime?* activity, it seems clear that students need help in providing evidence since they are not able to do so on their own. This difficulty for students to provide evidence is not only attributed to the lack of scientific content (Berland & McNeill, 2010) as stated in some papers, but to the connections that students have to make between such scientific knowledge and other factors that appear in the socio-scientific issues of daily life (Wu & Tsai, 2007), among them, beliefs about such issues, sometimes very far from scientific knowledge. Therefore, it is necessary to help students to be aware of their beliefs, of their origins (Zeidler, 1997) and to confront them with scientific knowledge about them.

Acknowledgements

This work is part of the ‘R&D Excelencia’ Project “Development of skills in problems of daily life through scientific practices of argumentation, inquiry and modelling in secondary and university education” (EDU2017-82197-P), funded by the Spanish Ministry of Economy and Finance through its 2017 research call.

References

- Andrée, M. (2005). Ways of Using “Everyday Life” in the Science Classroom. In K. Boersma, M. Goedhart, O. de Jong, & H. Eijkelhof (Eds.), *Research and the Quality of Science Education* (pp. 107–116). Dordrecht: Springer Netherlands.
- Barnes, L. L. B., & Wise, S. L. (1991). The Utility of a Modified One-Parameter IRT Model With Small Samples. *Applied Measurement in Education*, 4(2), 143–157.
- Bell, P., & Linn, M. C. (2000). Scientific arguments as learning artifacts: designing for learning from the web with KIE. *International Journal of Science Education*, 22(8), 797–817.

- 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., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education*, 94(5), 765–793.
- Bravo-Torija, B., & Jiménez-Aleixandre, M. P. (2018). Developing an Initial Learning Progression for the Use of Evidence in Decision-Making Contexts. *International Journal of Science and Mathematics Education*, 16(4), 619–638.
- Cebrián-Robles, D., Franco-Mariscal, A.-J., & Blanco-López, Á. (2018). Preservice Elementary Science Teachers' Argumentation Competence: Impact of a Training Programme. *Instructional Science*, 46(5), 789–817.
- Cebrián-Robles, D., & Rodríguez-Mora, F. (2018). Diseño de actividades de argumentación científica sobre el consumo de agua embotellada. Propuesta de evaluación mediante rúbricas [Design of scientific argumentation activities on the consumption of bottled water. Evaluation proposal through rubrics.]. In C. Martínez-Losada & S. García-Barros (Eds.), *28 Encuentros de Didáctica de las Ciencias Experimentales. Iluminando el cambio educativo* (pp. 637–642). A Coruña: Universidade da Coruña, Servizo de Publicacións.
- Deng, Y., & Wang, H. (2017). Research on evaluation of Chinese students' competence in written scientific argumentation in the context of chemistry. *Chemical Education Research and Practice*, 18(1), 127-150.
- Doria, M. F. (2006). Bottled water versus tap water: understanding consumers' preferences. *Journal of Water & Health*, 4 (2), 271-276.
- Doria, M. F. (2010). Factors influencing public perception of drinking water quality. *Water Policy*, 12(1), 1–19.
- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84(3), 287–312.
- Escobar-Pérez, J., & Cuervo-Martínez, Á. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances En Medicina*, 6(1), 27–36.
- Evagorou, M., & Avraamidou, L. (2008). Technology in support of argument construction in school science. *Educational Media International*, 45(1), 33–45.
- Evagorou, M., Jiménez-Aleixandre, M. P. & Osborne, J. (2012). 'Should we kill the grey squirrels?' A study exploring students' justifications and decision-making. *International Journal of Science Education*, 34(3),401–428.
- Ferrier, C. (2001). Bottled Water: Understanding a Social Phenomenon. *AMBIO: A Journal of the Human Environment*, 30(2), 118–119.
- Gavidia-Catalán, V. (2009). El agua en las enseñanzas mínimas de Secundaria [Water in

- the secondary teachings]. In *Agua y sostenibilidad: recursos, riesgos y remedios* (pp. 61–94). Ministerio de Educación Cultura y Deporte.
- Gleick, P. H., & Cooley, H. S. (2009). Energy implications of bottled water. *Environmental Research Letters: ERL [Web Site]*, 4(1), 014009–014015.
- Godden, D. M. (2010). The importance of belief in argumentation: belief, commitment and the effective resolution of a difference of opinion. *Synthese*, 172(3), 397-414.
- Gotwals, A. W., & Songer, N. B. (2013). Validity evidence for learning progression-based assessment items that fuse core disciplinary ideas and science practices. *Journal of Research in Science Teaching*, 50(5), 597-626.
- Grooms, J., Sampson, V. y Enderle, P. (2018). How concept familiarity and experience with scientific argumentation are related to the way groups participate in an episode of argumentation. *Journal of Research in Science Teaching*, 55(9), 1264-1286.
- He, X., Deng, Y., Yu, S., & Wang, H. (2020). The Influence of Context on the Large-Scale Assessment of High School Students' Epistemic Cognition of Scientific Argumentation. *Science & Education*, (29), 7-41.
- Henderson, J. B., McNeill, K. L., González-Howard, M., Close, K., & Evans, M. (2018). Key challenges and future directions for educational research on scientific argumentation. *Journal of Research in Science Teaching*, 55(1), 5–18.
- Hsieh, F.-P., Lin, H.-S., Liu, S.-C., & Tsai, C.-Y. (2019). Effect of Peer Coaching on Teachers' Practice and Their Students' Scientific Competencies. *Research in Science Education*, 1–24.
- Jiménez-Aleixandre, M. P., & Díaz-de-Bustamante, 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: Revista de Investigación Y Experiencias Didácticas*, 21(3), 359.
- Jiménez-Aleixandre, M. P., & Puig-Mauriz, 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 Didáctica de Las Ciencias Experimentales*, (63), 11–18.
- Jiménez-Aleixandre, M. P., Rodríguez, A. B., & Duschl, R. A. (2000). “Doing the lesson” or “doing science”: Argument in high school genetics. *Science Education*, 84(6), 757–792.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130-144.
- King, D., & Ritchie, S. M. (2012). Learning Science Through Real-World Contexts. In B. Fraser, K. Tobin, & J. C. MacRobbie (Eds.), *Second International Handbook of Science Education* (pp. 69–79). Springer Press.
- Lee, H.-S., Liu, O. L., Pallant, A., Roohr, K. C., Pryputniewicz, S., & Buck, Z. E.

- (2014). Assessment of uncertainty-infused scientific argumentation. *Journal of Research in Science Teaching*, 51(5), 581–605.
- Lu, J., & Zhang, Z. (2013). Assessing and Supporting Argumentation with Online Rubrics. *International Education Studies*, 6(7), 66.
- Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, 47(2), 149–174.
- 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. (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Brussels: OECD.
- OECD. (2012). *OECD science, technology and industry outlook 2012*. Paris: Organisation for Economic Co-operation and Development.
- Oliveras, B., Márquez, C., & Sanmartí, N. (2013). The Use of Newspaper Articles as a Tool to Develop Critical Thinking in Science Classes. *International Journal of Science Education*, 35(6), 885–905.
- Osborne, J., 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.
- Péronnet, F., Mignault, D., du Souich, P., Vergne, S., Le Bellego, L., Jimenez, L., & Rabasa-Lhoret, R. (2012). Pharmacokinetic analysis of absorption, distribution and disappearance of ingested water labeled with D2O in humans. *European Journal of Applied Physiology*, 112(6), 2213–2222.
- Pinochet, J. (2015). El modelo argumentativo de Toulmin y la educación en ciencias: una revisión argumentada [The argumentative model of Toulmin and the education in sciences: a review argued]. *Ciência & Educação (Bauru)*. Retrieved from <http://www.redalyc.org/pdf/2510/251038426004.pdf>
- Rasch, G. (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests* (Vol. 184). Oxford, England: Nielsen & Lydiche Studies in mathematical psychology.
- 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 [The “consumption of bottled drinking water” as a context for the development of scientific competences. A case study in the third year of compulsory secondary education] (Doctoral Thesis). Universidad de Málaga.
- Rodríguez-Mora, F., & Blanco-López, Á. (2012). Ideas y creencias de alumnos de educación secundaria sobre la presencia de cal en el agua de bebida [Ideas and beliefs of secondary school students about the presence of lime in drinking water]. In G. Pinto & M. Martín (Eds.), *Enseñanza y divulgación de la química y la física*

- (pp. 197–204). Madrid: Garceta Grupo Editorial.
- Sadler, T. D., & Zeidler, D. L. (2005). The significance of content knowledge for informal reasoning regarding socioscientific issues: Applying genetics knowledge to genetic engineering issues. *Science Education*, 89(1), 71–93.
- Stevens, D. D., & Levi, A. J. (2005). *Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback, and Promote Student Learning*. Virginia: Stylus Publishing, LLC.
- Simonneaux, L. (2008). Argumentation in socio-scientific contexts. In S. Erduran (Ed.), *Argumentation in science education: perspectives from classroom-based research* (pp. 179–199). Dordrecht: Springer.
- Siribunnam, S., Nuangchalerm, P., & Jansawang, N. (2014). Socio-Scientific Decision Making in the Science Classroom. *Online Submission*, 5(4), 1777–1782.
- Toulmin, S. E. (1958). *The uses of argument* (2003rd ed.). Cambridge: Cambridge University Press.
- Walton, D. N. (1989). Dialogue theory for critical thinking. *Argumentation*, 3(2), 169–184.
- WHO. (2011). Hardness in Drinking-water. Retrieved from http://www.who.int/water_sanitation_health/dwq/chemicals/hardness.pdf.
- Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago: MESA Press.
- Wu, Y., & Tsai, C. (2007). High School Students' Informal Reasoning on a Socio-scientific Issue: Qualitative and quantitative analyses. *International Journal of Science Education*, 29(9), 1163–1187.
- Zeidler, D. L. (1997). The central role of fallacious thinking in science education. *Science Education*, 81(4), 483–496.

Appendix

Table 1. Phases of the teaching sequence on the consumption of bottled water and questions addressed in each of them

PHASES OF THE SEQUENCE	QUESTIONS
<p><i>Phase I.</i> <i>Identification of the problem</i></p>	<ul style="list-style-type: none"> ▪ Why is it important to consider the consumption of bottled drinking water? ▪ What are the differences between bottled drinking water and tap water for you? ▪ What are your reasons for drinking bottled drinking water? ▪ Which of your reasons can be scientifically proven?
<p><i>Phase II.</i> <i>Treatment of the problem</i></p>	<ul style="list-style-type: none"> ▪ What is bottled water from a chemical point of view? ▪ What are the differences between tap water and bottled water? ▪ Why do we drink bottled water?
<p><i>Phase III.</i> <i>Synthesis and recapitulation</i></p>	<ul style="list-style-type: none"> ▪ What are the advantages and disadvantages of drinking tap water and bottled water? ▪ Do we have to drink bottled water? Is it better than tap water?
<p><i>Phase IV.</i> <i>Reflection on learning and assessment</i></p>	<ul style="list-style-type: none"> ▪ What have I learned about the subject?

Table 2. Hypotheses on the level of difficulty of assessment activities

Activity	Evidence that is provided by		Chemical knowledge	Level Difficulty
	The activity	The student		

description				
<i>Is bottled water pure?</i>	All	None	Teached	N1 - Low
<i>Does bottled water have lime?</i>	None	All	Teached	N2 - Medium
<i>Is high mountain water the same as tap water?</i>	None	All	No teached	N3 - High

Table 3. Frequency of students that reached a certain level of achievement in the elements of the argument

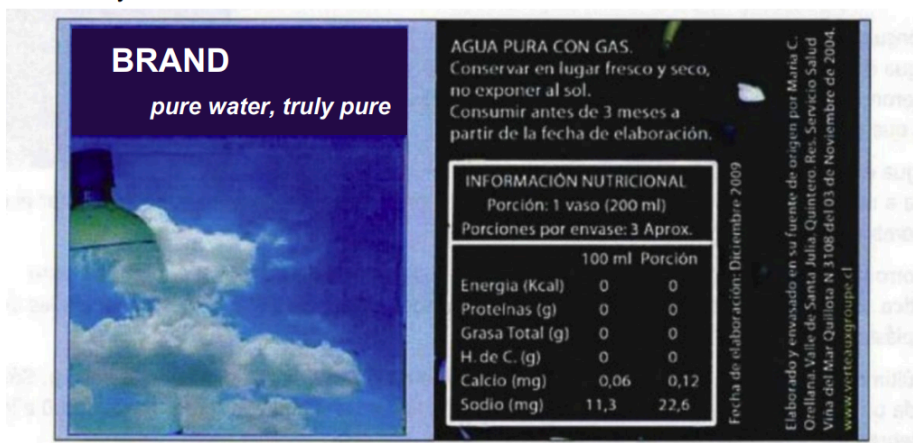
Activity	Total number of students that achieved the maximum value at different levels of the rubric														
	Evidence				Justification					Conclusion					
	1	2	3	4	1	2	3	4	5	1	2	3	4	5	
1. <i>Is bottled water pure?</i>	10	3	11	1	1	11	0	4	9	0	0	2	8	15	
2. <i>Does bottled water have lime?</i>	22	2	1	0	0	1	24	0	-	0	1	2	11	11	
3. <i>Is high mountain water the same as tap water?</i>	8	8	7	2	7	9	6	1	2	0	8	12	3	2	

Table 4. Frequency of students passing each of the elements of the argument

	Elements of the argument		
	Evidence	Justification	Conclusion
	1. <i>Is bottled water pure?</i>	12	13

2.	<i>Does bottled water have lime?</i>	1	24	22
3.	<i>Is high mountain water the same as tap water?</i>	9	9	6

A label on a bottle of a brand of bottled water (front and back). Read it carefully:



Do you think the water in this bottle is “pure water, truly pure” as the label says? Substantiate your answer based on the information (terms or data) that appears on the label.

Figure 1. Image of the *Is bottled water pure?* assessment activity.

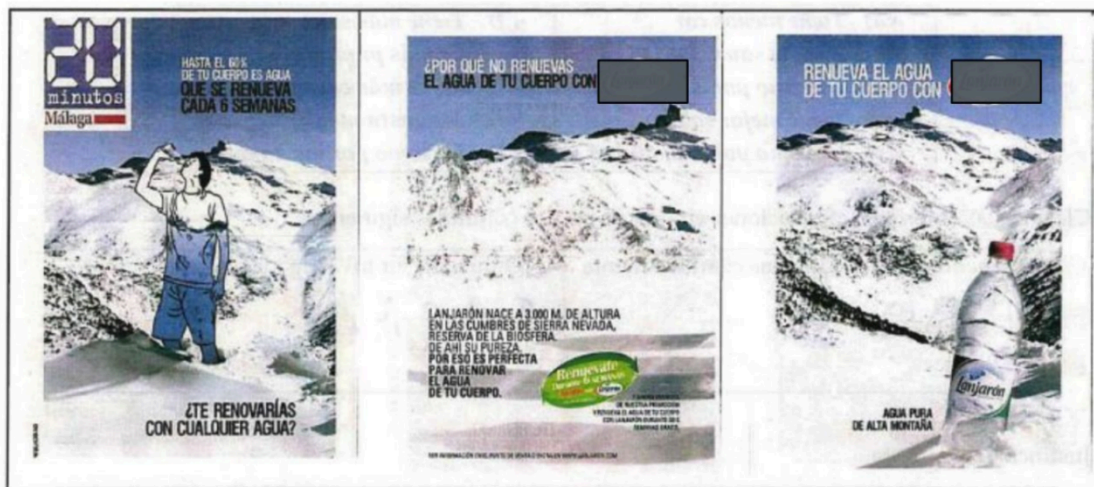
1. Evidence				
1	2	3	4	
There is no evidence	Although there is evidence, it has nothing to do with water purity	It provides some evidence that justifies the reason the water is not pure from a scientific point of view	Provides enough and appropriate evidence to defend that the water is not pure from a scientific point of view (which has gas, calcium and sodium)	
2. Justification				
1	2	3	4	5
There is no justification	There is a justification, but it does not link the evidence with the conclusions or does so without chemical knowledge	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	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	Provides a justification that links the conclusion to the evidence. There are sufficient correct scientific ideas with implicit chemical knowledge. Terms are used accurately without showing any doubt
3. Conclusion				
1	2	3	4	5
There is/are no conclusions (Unknown or it says it cannot be known)	Reaches an erroneous conclusion by thinking that the water is chemically pure	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	The conclusion is correct and well-defended scientifically but uses tentative terms (I think, in my opinion...)	The conclusion is correct and well-defended scientifically without showing any doubt

Figure 2. Specific rubric of the *Is bottled water pure?* assessment activity.

1.Evidence									
1	The student gives no evidence	2	The student gives evidence about bottled and tap water, but it is incorrect. The evidence is not appropriate and/or does not support the conclusion, for saying that it tastes better, for instance	3	The student provides evidence that both bottled water and tap water have mineral salts but does not give any examples of them	4	The student produces evidence that both bottled and tap water have mineral salts and provides some examples of them, such as calcium and magnesium		
2.Justification									
1	The student gives no justification whatsoever	2	The student makes some justification, but it has nothing to do with the statement or is mistaken by saying that bottled water is different from tap water in terms of whether mineral salts are present	3	The student makes a justification related to the presence of lime in tap and bottled water, but it is not enough to argue that there are mineral salts in the water that give it its hardness and that is why it is called lime	4	There is a justification related to the presence of lime in tap and bottled water and is sufficient to argue that there are mineral salts in the water that give it its hardness and that is why it is called lime		
3.Conclusion									
1	The student reaches no conclusion	2	The student reaches an inappropriate conclusion by saying that tap and/or bottled water does not contain lime	3	Despite knowing that both bottled and tap water contain lime, the conclusion is scientifically inaccurate by arguing that all water has lime	4	The student believes that both bottled and tap water contain lime, and the conclusion is scientifically accurate, but the formula he or she uses is tentative (I think, in my opinion...)	5	The student gives an accurate conclusion that both bottled and tap water contain lime. Besides, the conclusion is scientifically accurate without showing any doubt

Figure 3. Specific rubric of the *Does bottled water have lime?* activity.

Look at the following advertising statement and read it carefully:



Answer the following questions: Would you “renew yourself anyway” if instead of drinking “pure mountain water” you drank “tap water”? Explain your answer.

Figure 4. Image of the *Is high mountain water the same as tap water?* assessment activity.

1.Evidence				
1 The student gives no evidence	2 The student gives incorrect evidence about the presence of mineral salts in high mountain and tap water. These are not appropriate and/or do not support the conclusion	3 The student provides some evidence such as that mountain water or tap water have similar compounds but fails to explain the similar composition that makes it drinkable	4 The student provides sufficient and appropriate evidence to argue that high mountain and tap water have similar mineral salts that help renew the body equally	
2.Justification				
1 The student gives no justification whatsoever	2 The student makes some justification but thinks wrongly that high mountain water and tap water are different in terms of mineral salts, and how they can renew the body	3 The student makes a justification that states that both types of water can serve to renew the body but does not justify it scientifically based on salts but on assertions such as that one is safer or better tasting than the other	4 The student provides a justification that relates the conclusion to the evidence and includes appropriate and sufficient scientific ideas for the argument but does not use terms accurately and it is done showing doubts	5 The student provides a justification that relates the conclusion to the evidence and includes appropriate and sufficient scientific ideas for the argumentation. The student also uses the terms accurately and without showing any doubt
3.Conclusion				
1 The student reaches no conclusion	2 The student reaches an inappropriate conclusion by thinking that tap and/or high mountain water are different regarding their renewal capacity	3 Despite affirming that both renew equally, the conclusion is not very precise or does not allude correctly from a scientific point of view	4 The student thinks that high mountain water and tap water contain salts that also renew. He or she might be scientifically accurate but expresses him or herself tentatively (I believe, in my opinion...)	5 The student thinks that high mountain water and tap water contain salts that also renew. Besides, the conclusion is scientifically accurate and without any doubt

Figure 5. The rubric of the *Is high mountain water the same as tap water?* assessment activity.

Yo pienso que en el término científico no es puro porque el agua pura en química es aquella que solo está formada por una sustancia.

(I think that, in scientific terms, it is not pure because pure water in chemistry is formed only by one substance).

Figure 6. Student S05 response to the *Is bottled water pure?* assessment activity and coded as 1E2J4C

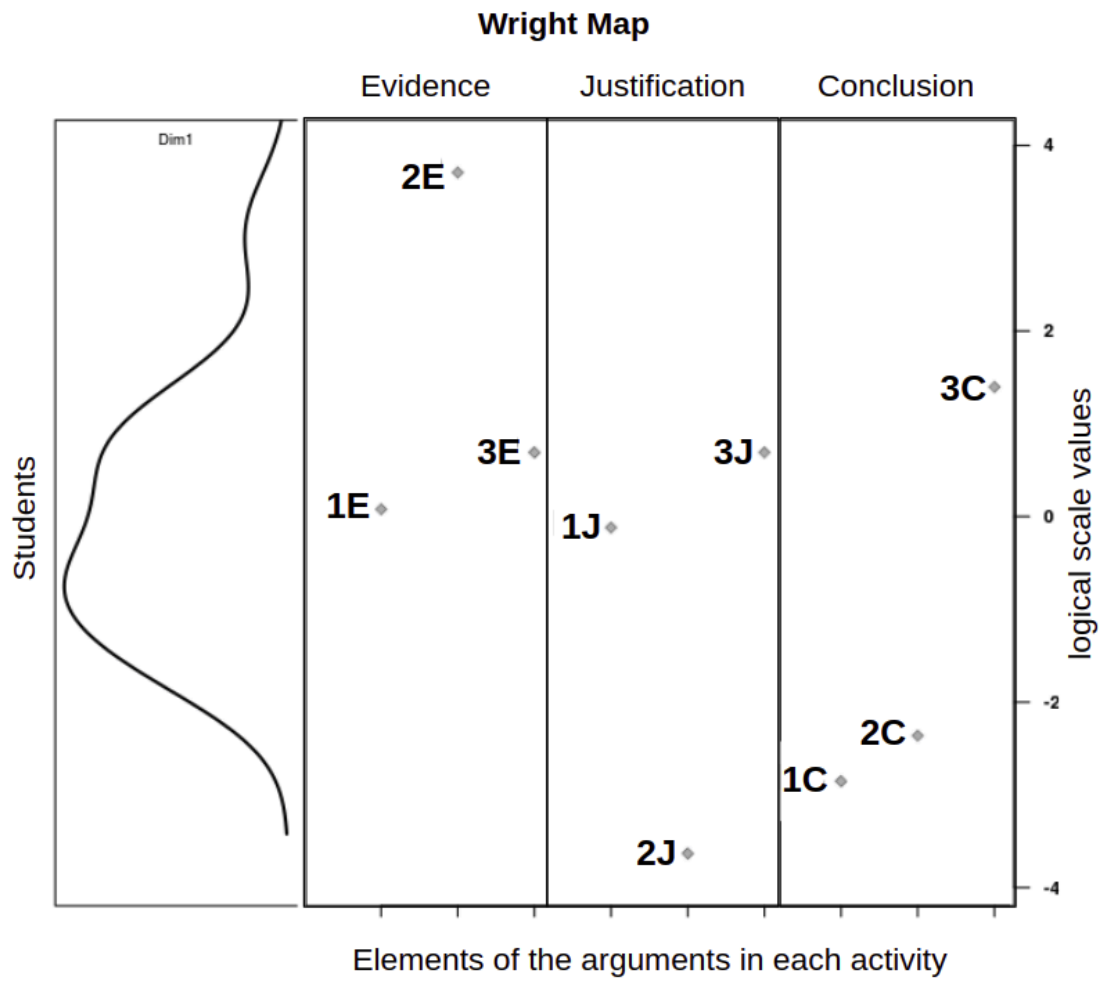


Figure 7. Wright's map showing the difficulty in reaching the levels considered adequate for each of the elements of the three arguments analysed.