



Psykhe

ISSN: 0717-0297

psykhe@uc.cl

Pontificia Universidad Católica de Chile  
Chile

Larrain, Antonia; Howe, Christine; Cerda, Julieta  
Argumentation in Whole-Class Teaching and Science Learning  
Psykhe, vol. 23, núm. 2, 2014, pp. 1-15  
Pontificia Universidad Católica de Chile  
Santiago, Chile

Available in: <http://www.redalyc.org/articulo.oa?id=96732496005>

- How to cite
- Complete issue
- More information about this article
- Journal's homepage in redalyc.org

redalyc.org

Scientific Information System  
Network of Scientific Journals from Latin America, the Caribbean, Spain and Portugal  
Non-profit academic project, developed under the open access initiative

# Argumentation in Whole-Class Teaching and Science Learning

## Argumentación en Enseñanza en Clase Completa y Aprendizaje de Ciencias

Antonia Larrain  
Universidad Alberto Hurtado

Christine Howe  
University of Cambridge

Julieta Cerda  
Universidad Alberto Hurtado

Peers' discussion of contradictory ideas has been proven to promote students' learning. Some empirical evidence suggests that whole-class argumentation has similar benefits, but there is no clarity yet on whether discussion accounts for this effect. This study aimed at testing the effects of different aspects of whole-class argumentation on science learning. A non-probabilistic sample of 220 students (aged 10 to 11 years) from 18 public schools in Santiago, Chile, participated in the study. Eleven teachers delivered lessons according to a teaching programme especially developed to foster argumentation (intervention group) and 7 teachers delivered lessons in their usual way (control group). Students were assessed individually using pre- and post-measures of learning, argumentative skills and attitudes toward science. The two formers were tests and the latter was a questionnaire. Lessons were videotaped. Factorial analysis and linear regression were conducted. Results showed that 2 factors predict a portion of the variance on learning: one factor composed of justificatory utterances and the other of students' counter-arguments. These results suggest that contradiction among peers is not the only aspect of classroom argumentation that prompts learning.

*Keywords:* argumentation, science learning, discussion, dialogic teaching, effective teaching

La discusión de ideas contrarias entre estudiantes promueve el aprendizaje en ciencias. Existe evidencia que sugiere que la argumentación en clase completa tendría el mismo efecto, pero no está claro aún que se deba a la discusión. Este estudio evaluó el efecto de distintos aspectos de la argumentación en clase completa en el aprendizaje de ciencias. En una muestra no probabilística participaron 220 estudiantes (10 y 11 años de edad) de 18 escuelas municipales de Santiago, Chile. Once profesores dieron clases diseñadas para promover argumentación (grupo intervención) y 7 profesores lo hicieron en su forma habitual (grupo control). Los estudiantes fueron evaluados individualmente antes y después con pruebas de aprendizaje y argumentación y cuestionario de actitud hacia la ciencia. Se grabaron las clases. A través de análisis factoriales y regresiones lineales, los resultados mostraron que 2 factores predicen una parte de la varianza del aprendizaje: uno compuesto por enunciados justificativos y otro, por contra-argumentos de los estudiantes. Estos resultados sugieren que la contradicción entre pares no es el único aspecto de la argumentación en clase completa que promueve el aprendizaje.

*Palabras clave:* argumentación, aprendizaje en ciencias, discusión, enseñanza dialógica, enseñanza efectiva

Many scholars have argued for the benefits of a pedagogical use of argumentation for science learning (Driver, Newton, & Osborne, 2000; Duschl & Osborne, 2002; Jiménez-Aleixandre & Erduran, 2008; Osborne, Erduran, & Simon, 2004). Although there is scarce research testing these benefits, some empirical evidence on peer interactions partially supports it (Asterhan & Schwarz, 2007, 2009; Howe et al., 2007). Some studies have tested the effect of the argumentative type of whole-class talk on learning, but they are only partially conclusive and do not prove the effect of the discussion of contradictory ideas (Che & She, 2012; Mercer, Dawes, Wegerif, & Sams, 2004; Venville & Dawson, 2010; Wilson, Taylor, Kowalski, & Carlson, 2010; Zohar & Nemet, 2002). Considering that whole-class interactions are the most frequent ones in classrooms around the world (Howe, 2010), it is relevant to investigate the effect of whole-class discussion on learning.

---

Antonia Larrain, Facultad de Psicología, Universidad Alberto Hurtado, Santiago, Chile; Christine Howe, Faculty of Education, University of Cambridge, United Kingdom; Julieta Cerda, Facultad de Psicología, Universidad Alberto Hurtado, Santiago, Chile.

The research reported was funded by the Chilean Fondo Nacional de Ciencia y Tecnología (FONDECYT), Project 11100181.

The authors are grateful for the generosity of the epiSTEMe team for giving their permission to use and adapt the project Forces materials: Kenneth Ruthven, Riikka Hofmann, Christine Howe, Stephanie Luthman, Neil Mercer, and Keith Taber. Also the authors want to thank the contribution of Muriel Vélazquez and Sulvy Cáceres.

Correspondence concerning this article should be addressed to Antonia Larrain, Facultad de Psicología, Universidad Alberto Hurtado, Avda. Almirante Barroso 26, Santiago, Chile. E-mail: alarrain@uahurtado.cl

A recent study conducted in a Chilean public middle school, whose aim was to examine the effect of oral argumentation on science learning, shows that whole-class argumentation is a predictor of science learning (Larrain, Howe, & Freire, 2014). Considering that two recent studies showed that whole-class discussions of contradictory ideas is scarce in Chilean public middle-school education (Larrain, Freire & Howe, 2014; Preiss, Larrain, & Valenzuela, 2011), it is relevant to explore whether it is the discussion or another component of argumentative discourse that has an effect on learning.

The goal of this paper is to report a study aimed at testing the effects of different components of whole-class argumentation on science learning. The analyses reported here were drawn from the corpus gathered as part of Project FONDECYT 11100181. The same corpus of data was used for others studies. On this occasion the idea was to test whether or not the effect of oral argumentation on learning can be attributed to the discussion of contradictory ideas.

### What Should Count as Argumentation?

As Jiménez-Aleixandre & Erduran (2008) point out, argumentation has different meanings. The empirical evidence mentioned above typically does not use the term *argumentation*, despite the fact that it clearly involves its use, in one way or another. For instance, the work of Howe et al. (2007) refers to the discussion of contradictory ideas, Mercer et al. (2004) use the term *exploratory talk*, referring to an explicit and collaborative style of reasoning, and Wilson et al. (2010) refer to *inquiry-based teaching*, as a kind of talk in which students engage, explore, explain, elaborate, and evaluate. All these notions partially overlap with what we refer to as argumentation.

Following the key ideas in some classical theories (Perelman & Olbrecht-Tyteca, 1958/1969; Toulmin, 1958; van Eemeren & Grootendorst, 1992), we consider argumentation to be the discursive practice in which speakers deal with controversial issues increasing their comprehension or acceptability. Although argumentation and explanation may coincide in the use of some discursive markers (e.g., Why? Because), in explanations one piece of discourse is offered as the efficient cause of a given fact or event (Osborne & Patterson, 2011) with the purpose of clarification (Asterhan & Schwarz, 2009). On the contrary, in argumentation one piece of discourse is offered in support of another piece of discourse that is recognised as relatively weak or controversial. There is a rhetorical context of virtual or real dispute rather than clarification. We conceive of rhetorical context as those aspects of discursive interactions which inform about its goals and type (hierarchy) of participants, constraining and defining the social setting.

Coherently, we conceive of argumentation as the discursive activity by which controversial standpoints are supported and critically evaluated, coordinating empirical and/or theoretical evidence with the goal of reaching an understanding. This understanding may involve only one speaker, who holds both perspectives (Billig, 1987; Leitão, 2000), one of which may also not have been explicitly formulated: the presence of justification suggests the existence of alternative positions and, in turn, the relative weakness of the supported position (see Toulmin, 1958). Consequently, justification is an indication of argumentation.

According to Leitão (2000), argumentation is a type of discourse which involves specific semiotic mechanisms that promote specific psychological processes of knowledge construction. These mechanisms are justification, counter-argumentation, and response. Whereas justification fosters awareness about the weakness of a position, thus forcing the speaker to support it, counter-argument prompts a shift of focus in attention from the object of discourse to the speaker's own thinking, promoting the revision of the grounds of one's opinions. Counter-argument is a key semiotic mechanism as a result of its role in fostering meta-cognitive processes that facilitate knowledge construction and critical thinking abilities. Hence, whereas explanation may be thought of as contributing to knowledge organisation (de Vries, Lund, & Baker, 2002), the controversial rhetorical context of argumentation prompts knowledge revision and elaboration.

Although theoretically any piece of argumentative discourse involves, in one way or another, the presence of contradiction and opposition, we will conceive of dialectic argumentation (Asterhan & Schwarz, 2009) when contradictory positions or ideas are explicitly formulated and supported. When there is no explicit discussion of contradictory ideas, but there are supported opinions that may be evaluated or revised without discussing opposite views, we will refer to as justificatory or one-sided argumentation. Consequently, discussion (of contradictory ideas) will be considered as only one form argumentation may take.

### Argumentation for Science Learning

Considering the psychological potential of argumentation (see Schwarz, 2009), its pedagogical use seems highly recommendable. In fact, empirical evidence clearly supports the effect of discussion of contradictory views among peers on science learning (Asterhan & Schwarz, 2009; Howe & Tolmie, 2003; Howe, Tolmie, & Rodgers, 1990, 1992; Howe et al., 2007). Moreover, Asterhan and Schwarz (2009) failed to find a relation between one-sided argumentation, or collaborative argumentation in favour of one idea, and science learning. Coherently, a study reported in Howe (2009) suggests that contradictions that were not resolved during peer interaction were associated with greater conceptual gains. On the contrary, joint constructions were not clearly associated with progress. In these studies, the effects of discussion on science learning were greater in tests that were taken weeks after peer interactions (delayed post-tests).

Considering the effect of peer argumentation, it is worth noting that no studies have properly proved the effect of whole-class discussion on science learning. Consequently, although there are good theoretical reasons to think that whole-class discussions have a positive impact on science learning, there is no conclusive evidence to support this belief. The empirical evidence available on the effect of whole-class argumentation does not demonstrate the effect of discussion in particular. Results from the study reported in Mercer et al. (2004) suggest that exploratory talk, or the discursive practice in which students express and collaboratively explore their opinions through a systematic and joint reasoning, has an impact on secondary students' science learning and reasoning skills. In addition, results from four studies with primary (Che & She, 2012) and secondary students (Venville & Dawson, 2010; Wilson et al., 2010; Zohar & Nemet, 2002) support this conclusion. However, there are uncertainties about how these results should be interpreted.

First, these studies control neither the frequency and type of *argumentative* discourse nor the social organisation in which it occurs (among peers or whole-class interaction). Conclusions are drawn from group mean comparisons (intervention versus control). It is important to consider that, according to Sun, Bradley, and Akers (2012), gender, socio economic status (SES), motivation, and self-efficacy, among others, account for students' achievements. Second, even when the effect may be attributed to classroom argumentative talk, the studies do not enable conclusions to be drawn about the differential effects of discussions of contradictory ideas versus other forms of argumentation.

The first point was tackled by a recent study reported in Larrain, Howe, et al. (2014), which was conducted with Chilean students who attended public middle schools. The study evaluated the effect of whole-class argumentation on physics learning. The results showed that whole-class argumentation is a predictor of delayed science learning gains (post-test taken approximately one month after teaching), when controlling for school, condition, small peer-group argumentation and initial students' measures on learning, individual argumentation skills, and attitudes towards science. Nevertheless, once again, this study does not inform about the effects of different aspects of argumentation. It is not possible to ascertain whether the effect of whole-class argumentation on learning is attributable to discussion or other form of argumentation. Although the more conclusive evidence on the impact of argumentation on science learning is the one on peer discussions on contradictory views (Asterhan & Schwarz, 2009; Howe, 2009), recent empirical evidence on Chilean public middle-school whole-class science teaching shows that dialectic argumentation is extremely infrequent. Hence, it is possible that the effect reported in Larrain Howe, et al. (2014) is associated with other forms of argumentation. If that were the case, the evidence would be relevant, because it would suggest that learning benefits from reasoning as a whole and not just from socio-cognitive conflicts.

### Research Goals

The goal of the study was to evaluate the effects of different aspects of oral whole-class argumentation on science learning. We sought to determine whether the discussion of contrary ideas was the factor that predicts learning gains. Given the previous discussion, we expected to find an effect of discussion of contrary ideas on learning gains.

## Method

The whole-classroom interactions that we analysed in this study were drawn from a corpus of data gathered as part of the aforementioned experimental study (Project FONDECYT 11100181) on the effects of oral classroom argumentation on science learning (Larrain, Howe, et al., 2014). Given that the data collection was the same for both studies, the description of the rationale of the study may be very similar to that reported in Larrain, Howe, et al. (2014).

### Participants

We conducted a non-probabilistic sample and invited to participate every public school within a vulnerable area of Región Metropolitana (RM), Santiago, Chile, composed by five administrative counties. A total of 18 teachers (11 females) and fifth-grade classes (11 as the intervention group) from 18 public schools from RM participated in the study. Schools were equivalent in terms of school size and parents' SES (medium-low). On average teachers were 46 years of age. We invited every school from the mentioned five counties, through the school principals, to participate in the study. Although initially 28 schools agreed to participate and were assigned to two conditions (intervention and control groups), 10 teachers (3 intervention and 7 controls) left the study during its first phase. The main reason for this was that teachers were overloaded with other school commitments and projects. We invited all fifth-grade students from the 18 schools to participate. From a total of 538 students registered in the relevant classes, 220 students gave their own and their parents' written permissions (109 females). Classes were randomly assigned to both intervention (137 students, 73 females) and comparison groups (83 students, 36 females).

### Design and Procedure

The study was quasi-experimental, with two conditions (experimental and control groups) and pre-post measures. Intervention teachers were asked to deliver lessons on physics according to a teaching programme especially developed by the research team to foster argumentation (Forces). Control teachers were asked to deliver their lessons in the usual way. Teachers from both groups were asked to obtain student' measures (learning, individual argumentation skills, and attitudes questionnaire) both before and after teaching the unit on Forces. Teachers in the intervention group were also asked to develop Forces lessons (10) according to the project materials.

The first phase of the study consisted of collecting the permissions of teachers, parents and children. All participants were given full information about the project. Participant teachers attended a full training day at the university. Intervention group teachers received the lesson materials. Then, teachers took pre-test measures during science lessons. According to the Chilean national curriculum, Forces should have taken 10 lessons. On average, teachers delivered 13.3 lessons on the topic. No differences between groups were observed,  $F(1, 108) = 1.65$ ,  $p = 0.201$ ). Two lessons from each class (intervention and control groups) were videotaped. In classes where small-group work took place, we also made recordings of three groups. After finishing the unit, teachers took post-measures (learning and argumentation skills) on two occasions: one immediately after finishing the unit and the other four weeks later, on average.

### Classroom Materials

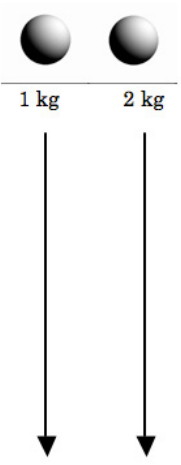

Some of the project materials were adapted from a parallel project conducted by a team based in the Faculty of Education, University of Cambridge, United Kingdom (Ruthven et al., 2011). The project, called epiSTEMe, sought to design and evaluate modules that include effective pedagogical principles that are relevant for the early years of British secondary education (12-13 years). The Cambridge team developed classroom materials that covered dialogic principles for classroom interaction, real-life examples, and practical hands-on work relating to four topics in mathematics and science. We focused on one of the science topics, Forces, the material which is relevant for the Chilean fifth grade (10-11 years), that is, a little earlier than in the United Kingdom.

We adapted the Forces module from epiSTEMe (Howe et al., 2014) in collaboration with a science teacher. The adaptation involved the re-design of several lessons (five) insofar as the British and Chilean curricula on Forces differ. Activities were carefully designed following expert recommendations (Andriessen & Schwarz, 2009; Leitão, 2009) to promote argumentation. Two examples of activities are presented in Table 1 and

Figure 1. The final module consisted of nine lessons on Forces and Movement. After the project finished, we also gave these materials to control group teachers.

Table 1

*Example of Whole-Class Activity: Intervention Group*

Type of activity	Description
Whole class interaction	<ol style="list-style-type: none"> <li>1. The teacher asks students to look at the picture. Assuming there is no friction, the teacher should ask: Which of the spheres falls first? Why?</li> <li>2. First, students think about the problem individually and write down their answers.</li> </ol> <p><b>PICTURE 10:</b></p> <div style="display: flex; align-items: center; justify-content: center;"> <div style="text-align: center; margin-right: 20px;">  </div>  </div>
Individual	
Small groups	<ol style="list-style-type: none"> <li>3. Then, they discuss their ideas within their group. The teacher asks them to think whether they change their mind when talking to the group. Groups reach an agreement.</li> <li>4. Finally, they share their answers with the rest of the class.</li> </ol>
Whole class interaction	<ol style="list-style-type: none"> <li>5. The teacher provides 3 alternatives which allow students to reach the answer. He/She asks each group to choose an alternative according to their discussion. Students discuss the alternatives and choose one, which they will work on again at the end of the lesson:             <ol style="list-style-type: none"> <li>a) The lighter sphere reaches the ground first, because it weighs less than the sphere with more mass.</li> <li>b) The sphere with more mass falls first, because it weighs more and speed depends on mass.</li> <li>c) Both spheres fall at the same time, because the acceleration of the fall is not dependent on mass.</li> </ol> </li> </ol> <p>Afterwards, the teacher states that C is the correct alternative.</p>




## Accelerated car

**Read through each statement. On your own mark each of the statements ✓ or ✗ according to whether it is right or wrong.**

**Consider each of the statements together, and provide reasons why you think each is right or wrong.**

**Come to a group decision about which statement is true and write down the letter in the box.**



**A** There is only one force – the thrust from the engine. That's why the car is accelerating.

**B** The thrust from the engine is bigger than the friction and the air resistance together. That's why the car is accelerating.

**C** The air cannot push. It is just there and does not affect the speed of the car.

© epiSTEMe 2009/10

Figure 1. Example of small group activity: Intervention group.

## Measures

### Students' measures.

**Science learning.** During the first year of the project we adapted the knowledge tests developed for epiSTEMe (Howe et al., 2014). An independent judge rated the test in term of its validity in relation to the learning goals it was supposed to measure. We conducted a pilot study in which 239 fifth-grade students from nine Chilean public schools participated in answering the three versions of the test (pre, post-immediate and post-delayed). We obtained three versions equivalent in terms of item difficulty and internal consistency. The Cronbach alpha coefficients for each version were: pre-test  $\alpha = 0.89$ , post-immediate  $\alpha = 0.92$ , and delayed post-test  $\alpha = 0.89$ . Each version included 24 items on the following topics: movement, trajectory, displacement, speed, force, balanced forces, weight and mass, and types of force. Nine questions measured conceptual understanding, 12 included conceptual application to real-life examples, and three assessed scientific thinking skills. The total score of the test was obtained by adding each question score (32 points).

**Individual argumentation skills.** In order to control the effect of students' initial individual argumentation skills on learning, we took a written test. During the first year of the project we developed a written argumentation skills test. The pilot study and final instrument are reported in Larrain, Freire, and Olivos (2014). An independent judge rated the test in term of its validity in relation to the learning goals it was supposed to measure. Two trained judges independently coded 30% of the tests. Cohen's Kappa scores were acceptable for all questions: four items were perfect,  $K = 1$ , two were excellent,  $K > 0.84$ , four were very good,  $K > 0.78$ , and the remaining three were acceptable,  $K > 0.5$ . The total score for the test was obtained by adding each question score (28 points).

**Attitudes towards science.** In order to control the effect of motivational variables on learning, we included a questionnaire on attitude towards science. Again, during the first year of the project we adapted epiSTEMe's attitudes towards science tests (Howe et al., 2014). We conducted a pilot study in which 175 fifth-grade students from nine Chilean public schools participated in answering the test (one version). The overall internal consistency was very good,  $\alpha = 0.92$ . The adapted questionnaire was composed of 20 items.

Items were organised according to five aspects: disposition towards science study, disposition towards deep learning, valorisation of science study, perspective on own capability for science study, and perspective on own engagement with science.

### Classroom measures.

**Whole-class videos.** In order to assess the amount of argumentation used in lessons, two sessions of each class were videotaped. We used a coding scheme originally developed to analyse argumentative utterances in science classroom talk (Larrain, Freire, & Howe, 2014), identifying: (a) speaker, (b) semiotic mechanism (justifications or reasons, objections and counter-arguments), (c) argumentative questions (*Does anybody disagree? Does anybody think something different?*), and (d) explicit formulations of controversies. These codes are defined and exemplified in the excerpt presented in Table 2. Two trained judges coded independently 30% of the videos, using *The Observer XT Noldus®* (Jansen, Wiertz, Meyer, & Noldus, 2003). However, Cohen's Kappa scores on average were only  $K = 0.5$ . We decided to independently double-code the rest of the videos. All disagreements were discussed and resolved.

**Small-group register.** When videotaped lessons included small-group work, we recorded three group interactions per session (one video and two audio). We analysed the group discussions in order to control the effect of these discussions on learning. Although whole class and group interactions are complexly intertwined and it is not possible to clearly attribute learning to one of them separately, we looked for controlling the variance due to small group talk over whole class teaching. We developed a coding scheme to analyse small-group work argumentation based on the codes defined and exemplified in Table 2. Two trained judges coded 50% of the transcripts. Differences were discussed and resolved. The Cohen's Kappa scores for all utterances were acceptable: four codes were perfect,  $K = 1$ , two were excellent,  $K = 0.81$  and  $K = 0.91$ , four were good,  $K = 0.66$ , and one was only acceptable,  $K = 0.50$ . Once coders had reached agreement, one of them coded the remaining material. We calculated the total score for each observation corresponding to the frequency of argumentative utterances observed.

### Analyses of Data

Given the structure of the data and in order to meet the study goals controlling class-level variance and strengthening the parameter estimations, it would have been desirable to run multi-level model regressions. Although authors, such as Gelman and Hill (2007), indicate that such models can be run using a small sample size, the literature in general does not advise its usage for less than 30 groups (Bell et al., 2010, April; Maas & Hox, 2005; Scherbaum & Ferreter, 2009). We had only 18 classes at the group level, which is the relevant level for our estimations (Snijders, 2005). Moreover, we had only 10 classes and 6.6 students on average in each class with the relevant measures. Consequently, we used linear regression, including schools and groups as control variables. We also run an analysis of variance (ANOVA) in order to test the statistical significance of some group differences.

Considering that in the study of Larrain, Howe, et al. (2014) whole-class total frequency of argumentative utterances was a significant predictor only of delayed learning gains (when factors, such as school, condition, initial measures and small-group argumentation, were controlled), and that in this study we wanted to explore which components of this whole-class argumentation may explain those gains, the analyses are restricted to delayed learning gains controlling for the same variables. We chose delayed learning gains instead of delayed learning scores, because we wanted to eventually compare our results to epiSTEMe results (Howe et al., 2014), and also because we were interested in accounting for gains and not merely scores.

In order to sketch the different components of the argumentative discourse involved in the interactions that we analysed, we ran an exploratory principal component analysis with varimax rotation on the nine utterance variables for the whole sample. We decided to run this analysis because we wanted to know whether argumentative utterances' use was structured by interactive patterns distinguished by the presence of explicit contradiction and participants, as was observed in the study of Larrain, Freire, and Howe (2014).



Table 2  
*Definitions and Examples of Whole-Class Argumentation*

Participant	Utterances	Codes	Definitions
T	Where will the car move faster, Yaritza, on the smooth surface or on the corrugated one?		
S1	On the smooth one.		
T	Okay, the million dollar question, why?	Justificative question	Ask for reasons and justifications.
S4	Because there is less friction.	Justification	To formulate reason/s supporting a claim.
T	Miguel, Gianfranco, why? Let's learn to listen, where does it move more slowly?	Justificative question	
S5	On the corrugated paper.		
T	Why?	Justificative question	
S6	Not on the smooth surface.	Counter-position	Claim that contradicts a previous claim.
T	But let's try to reach an agreement, on which of the two...		
S6	On the smooth surface because on the other the wheels have more traction.	Counter-argument	To formulate reason/s supporting a counter-position.
T	So you disagree with the previous group; you say that the car will move faster on the corrugated surface than on the smooth one.	Formulation of controversy	To formulate explicitly an existing controversy.
T	What about this group over here? Where does it move faster?		
S1	On the smooth one.		
T	Why?	Justificative question	
S1	Because its gravity on the table makes it move better.	Justification	
	Okay. Miguel, do you agree with his answer?	Argumentative question	Question that invites people to agree or disagree with a given claim.
SS	No, no.		
	He said that the car would move faster on the corrugated surface than on the smooth one. Does anybody think Gianfranco is wrong?	Argumentative question	
SS	Yes, yes.		
T	Why?	Justificative question	
S4	Because the corrugated paper generates friction and lowers the car's speed.	Justification	

Note. T = teacher; S = student; SS = students

To test whether the factors had an effect on delayed learning gains, we conducted four multiple regression analyses using delayed learning gains as the dependent variable. Our interest was to see whether the interactions identified (and not each isolated code) predicted learning gains.

It is worth noting that although 220 students gave their permission to participate in the study, not all of them had all the measures. Some teachers did not take all measures and, in some cases, students were absent at the time measures were taken.

## Results

ANOVA tests by gender revealed no differences between boys and girls on immediate,  $F(1, 124) = 0.14$ ,  $p = 0.710$ , and deferred learning gains,  $F(1, 65) = 0.69$ ,  $p = 0.410$ , and gender was therefore not included in any of the models. Table 3 shows the means and standard deviations of the total sample for all measures

by condition. No differences between the two conditions were found at the beginning of the study with two control measures, science learning,  $F(1, 176) = 2.08, p = 0.150$ , and attitudes towards science,  $F(1, 164) = 1.68, p = 0.197$ , but there was a difference in individual argumentation skills,  $F(1, 188) = 8.86, p = 0.002, \eta^2 = 0.05$ , 95% IC [1.02, 4.45], having the intervention students initially better individual argumentation skills. The intervention group showed more learning gains than the control group in the delayed tests, which is similar to epiSTEMe results (Howe et al., 2014). However, unlike epiSTEMe, differences in delayed learning gains among groups were not significant,  $F(1, 65) = 0.61, p = 0.439$ .

Table 3  
*Descriptive Statistics of Variables per Condition*

Variable	Intervention			Control			Total		
	<i>N</i>	Mean	<i>SD</i>	<i>N</i>	Mean	<i>SD</i>	<i>N</i>	Mean	<i>SD</i>
Learning pre test (Lpre)	120	9.00	4.75	59	8.03	3.31	179	8.68	4.35
Learning delayed post test (LpostD)	47	13.81	5.55	25	9.28	4.41	72	12.24	5.59
Learning gains pre-post delayed (LG)	44	4.00	5.61	22	2.86	5.54	66	3.62	5.57
Individual argumentation pre test (IApre)	118	17.03	5.55	72	14.43	6.26	190	16.04	5.95
Attitudes towards science questionnaire (ATSQ)	109	95.23	15.92	58	91.93	19.77	167	94.08	17.36
Small-group work argumentation (SGAR)		3.96	4.46		1.83	1.45		3.17	3.78
Whole-class total number of argumentative utterances		14.28	0.81		18.25	1.03		16.26	0.66
Teachers' (T) justificatory questions (TJQ)		5.74	5.69		5.59	3.31		5.80	4.90
T justifications (TJ)		2.26	1.59		4.54	4.42		3.11	3.27
Students' (S) justifications (SJ)		6.40	6.86		7.33	4.20		6.75	6.09
S counter-positions (SC)		0.13	0.33		0.28	0.49		0.19	0.23
T counter-positions (TC)		0.07	0.26		0.24	0.42		0.14	0.34
T argumentative questions (TargQ)		1.06	1.81		0.48	0.79		0.84	1.55
T explicit formulation of controversy (TeC)		0.16	0.46		0.00	0.00		0.10	0.36
T counter-arguments (Tcount)		0.02	0.10		0.00	0.00		0.01	0.08
S counter-arguments (Tcount)		0.05	0.14		0.06	0.16		0.05	0.18

### Does Whole-Class Discussion of Contradictory Ideas Account for Learning Gains?

Table 4 shows the solution after the rotation in factor analysis. There were four factors with eigenvalues greater than one, and these factors jointly explained 76.8% of the total variance on observed utterances. The first factor involves Teacher justification questions (TJQ), Student justifications (SJ) and Teacher explicit formulation of controversies (TeC). Given that the frequency of TeC was so low in comparison with the other utterances, we called this factor Teacher-Students Justificatory Interaction. The second factor involves Teacher counter-positions (TC) and Argumentative questions (TargQ). We called this factor Teacher's Dialectic Interventions. The third factor involves Student counter-opinions (SC) and Counter-arguments (Scount). We called this factor Students Discussions. The fourth factor involves Teacher justifications (TJ) and Counter-arguments (Tcount). We named this factor Teacher Reasoning.

Table 4  
*Rotated Component Matrix for Utterance Codes*

Utterance	Component			
	1	2	3	4
	(eigenvalue = 2.8, 31.1% of variance)*	(eigenvalue = 1.46, 16.2% of variance)*	(eigenvalue = 1.32, 14.7% of variance)*	(eigenvalue = 1.32, 14.6% of variance)*
TJQ	0.901	0.116	-0.075	0.103
TJ	-0.219	0.105	0.245	0.808
SJ	0.900	0.124	-0.060	0.076
SC	-0.018	0.029	0.809	0.047
TC	-0.110	0.822	0.221	0.373
TargQ	0.261	0.856	-0.160	-0.238
TeC	0.791	-0.090	0.306	-0.256
Tcount	-0.291	0.060	0.247	-0.623
Scount	0.588	-0.023	0.610	0.046

\* PCA solution: Rotation sums of squared loadings.

Extraction method: principal component analysis; rotation method: Varimax with Kaiser normalisation. Rotation converged in five iterations.

Regarding the observed frequency of factors, it is interesting to note that we detected Factor 1 in 100% of the observed classes; in 50% of the groups this factor was observed on average between two and 11 times; in 27.7% of the groups, on average between 12 and 16 times; and in the rest of the groups, between 20 and 30 times. On average, Factor 1 occurred 12.7 times per lesson ( $SD = 7.82$ ). Factor 2 was noted in 83.3% of the observed classes; in 50% of the groups this factor was observed only once; and in the rest of the groups, between two and six times. On average, Factor 2 occurred 1.4 times per lesson ( $SD = 0.98$ ). Factor 3 was found in only 61.2% of the observed classes between one and two times. On average, Factor 3 occurred 0.5 times per lesson ( $SD = 0.49$ ). Finally, Factor 4 was observed in 100% of the classes; in 55.5% of the groups it was observed between one and two times; in 27.7% of the groups, between three and five times; and in the rest, between six and 17 times. On average, Factor 4 occurred 3.1 times per lesson ( $SD = 3.04$ ).

In multiple regression analyses, predictors were six control variables (school, condition and initial levels of learning, attitude towards science, written argumentation skills, and small-group frequency of argumentation) and the four factors identified above as independent variables. Only Factor 1 (Teacher-Students Justificatory Interaction) and 3 (Students Discussions) were predictors of learning gains. The model shown in Table 5 explains 58.5% of the variance in delayed learning gains,  $F(7, 23) = 7.03$ ,  $p < 0.001$ . The model shown in Table 6 explains 61.6% of the variance in delayed learning gains,  $F(7, 23) = 7.87$ ,  $p < 0.001$ .

We observed interesting results. Discussion of contrary ideas, as the discursive practice in which not only these ideas are expressed but also supported (see example of Table 3), would foster learning also in whole class interactions. Moreover, not only explicit discussion of contradictory views accounts for learning. Justificatory interactions also predicted learning gains. In both cases, condition and initial measures of learning are also predictors of learning gains, as could be expected. In fact, in epiSTEMe analyses initial measures of learning were also significant predictors of learning progress in both conditions (Howe et al., 2014). In the second regression, also school and small group argumentation were significant predictors. In the latter, this may suggest that when only students' argumentative discussions in whole class teaching are taken into account, there is a portion of the variance on delayed learning gains that is accounted for other patterns of discursive interactions. Since our analyses were very limited, due to the small simple size, in the following section we will take a closer look at some excerpts in order to comprehend this effect.

Table 5  
*Multiple Regression Coefficients for Delayed Learning Gains With Control Variables and Factor 1 as Predictor*

Variable	<i>B</i> (Standardised)	<i>T</i>
School	0.21	1.01
Condition	-0.60	-3.53*
IApre	0.20	1.17
Lpre-	-0.51	-3.50*
ATQ	0.08	0.63
SGAR	-0.03	-0.13
Factor 1	0.63	3.13*
Teacher-Students Justificatory Interaction		

\*  $p < 0.01$

Table 6  
*Multiple Regression Coefficients for Delayed Learning Gains With Control Variables and Factor 3 as Predictor*

Variable	<i>B</i> (Standardised)	<i>T</i>
School	1.49	4.10*
Condition	-1.93	-4.02*
IApre	0.19	1.23
Lpre	-0.44	-3.13*
ATSQ	0.01	0.14
SGAR	-1.56	-2.96*
Factor 3	2.48	3.53*
Students Discussions		

\*  $p < 0.01$

### Whole-Class Justificatory Interactions and Learning

Our analysis covered all teacher-students justificatory interactions (Factor 1). Although they vary in length, density of concepts, and depth of elaboration, in classes wherein the frequency of Factor 1 was greater than 15 (6 classes), these episodes were characterised by careful use of counterfactual information. The following excerpt corresponds to one of the classes in which students achieved higher learning gains:

- 1 Teacher (T): First, let's start playing with the ball. Let's throw the ball among ourselves. So? Let's throw it. What happens?  
 [Children start throwing the ball.]
- 2 T: Stop, the game is over. I saw the ball moving. Why did it move?
- 3 Student (S): Because we apply force when we throw the ball.
- 4 T: But how? I did not see that. Where did you get your force from? From your arm? Where do you get your force?
- 5 S1: We got the force when we are in contact with an object.
- 6 T: I'm not an object, I'm a person. Asael, okay, I'll let you take my arms; you pull me this way and you that way.  
 [The children pull the teacher towards them.]
- 7 T: What's happening? Am I moving? Is there a movement? Am I moving too much?  
 [The children pull the teacher from both sides.]
- 8 S2: No.
- 9 S3: Yes.

- 10 T: What would happen if John were stronger or if he apply more force than Asael? What would happen to me? Would I move?
- 11 S4: You'd move that way.
- 12 T: I'd fall, then. You can be sure I'd end up on the floor. And, why didn't that happen to me? What would happen?
- 13 S3: Because they had the same strength.
- 14 T: They applied the same force.

The teacher initiates the discursive interaction about the ball's movement by asking a *Why* question. One student gave an answer that might be conceived of as an explanation for the efficient cause of the ball's movement. However, the intervention in turn 4 transforms the rhetorical context, forcing the student to think about the evidence in which he is supporting his opinion. Although it is not clear whether it is a counter-argument or a contrary idea, it is an utterance that clearly opens up the possibility of alternative explanations. The interesting fact is that the student's answer was right; however, the teacher, instead of evaluating it and closing the interaction, introduces ambiguity. Furthermore, in turn 6 the teacher, instead of accepting the student's new answer, introduces a new (counterfactual) scenario, thus forcing the student to think about the concept he is using and raising the intellectual challenge. This counterfactual register is maintained in the next two teacher interventions. In fact, the second teacher's *Why* question draws on this new counterfactual scenario, which turns out to be a very effective tool for students' thinking promotion. It is interesting to note that although in the excerpt there is opposition and contradiction, contradictory ideas are not explicitly formulated and supported. Hence, we did not conceive the interaction as a discussion, but as a collaborative process of reasoning intending to reach a shared idea or conclusion without exploring contradictory ideas. Now take, for instance, the next excerpt:

- 1 T: We said our solar system is made up by a large star called the sun. This means we're only speaking about a star called the sun and not about the stars that surround it. Now, if you look at the sky at night, you see thousands of other stars. Why do you believe those stars look that way and not like our sun? Maybe they're different from the sun.
- 2 S1: Yes.
- 3 T: Why are they different?
- 4 S2: Some are bigger and others are smaller.
- 5 T: Do you think it may be because the sun is bigger? Ronny?
- 6 S3: The sun shines more and the stars shine less.
- 7 T: Why do the others shine little?
- 8 S4: Because they're older.
- 9 S5: Because some are smaller and the others are bigger.
- 10 T: So, could it be the size, Oriana?
- 11 S6: Because the stars are like fireballs.
- 12 T: And what's the sun?
- 13 S2: A star.
- 14 T: A star; so what are the other stars like?
- 15 S2: They're smaller.
- 16 T: But are they like fireballs too?
- 17 Students (SS): Nooo.
- 18 T: Why not?
- 19 S6: Yes.
- 20 T: Let me ask you, what's this book made of?
- 21 SS: Paper.
- 22 T: And this book?
- 23 SS: Paper.
- 24 T: So, they can vary in size and colour, but they are made of the same stuff. So, if I tell you that books are made of paper, what are stars made of?
- 25 S6: Fire.
- 26 T: Okay, let's suppose it's fire. What about the other stars?
- 27 S6: Fire.
- 28 T: And why is that?
- 29 S7: Because they're the same.
- 30 T: Because they're the same. If I take the smallest star and I compare it with the sun, the composition is the same. Now, what is the difference? Why do I see the sun like this, but the other stars do not look like it?
- 31 S3: Because the sun is closer and the others are farther away.



Again the teacher initiates the interaction with a *Why* question. This time, however, the question was about the cause not of a physical fact but of a psychological one (*Why do you believe?*). Moreover, the question is built on a problematic scenario that has a conditional structure. In turn 5 also the teacher, instead of evaluating and closing the interaction, poses a new counterfactual question using mental words. This new question opens up a new scenario in which other explanations are possible and necessary. This counterfactual and conditional register is maintained throughout the next teacher's interventions, forcing the students to think of their own answers, evaluate and elaborate them and formulate new explanations.

The two excerpts presented above shed light on the potential of different justifications for knowledge construction. Although these kinds of episode cannot be considered clearly as class discussions of contradictory ideas, they configure, and in turn are configured by, a virtual space in which alternative opinions and explanations operate, pushing students' thinking and ideas forward. They are clearly episodes in which collaborative reasoning occurs. However, we cannot say that it is a collaboration that is divested of all tension and social conflict. Students are made tense by teachers' counter-questions and counterfactual elicitations and it is probably this tense virtual field of ideas opened by the teacher that is a key aspect of the potential of these episodes for knowledge construction.

### Discussion

These findings shed light on the effect of argumentation on scientific learning beyond explicit conflict. It contributes to the growing body of evidence supporting the relevance of promoting the pedagogical use of argumentation in science learning. Such evidence should orient teachers' and school's efforts to transform classroom talk in reasoning and thinking spaces. Moreover, and following concerns raised by Howe and Abedin (2013), these results contribute to our knowledge of the effects of different forms of classroom talk.

In particular, the results support the idea that the discussion of contradictory ideas is beneficial for science learning, showing that this is not only the case when the discussion takes place among peers. From our results, discussions moderated and fostered by teachers in whole-class interaction are also effective even when they have different characteristics and may prompt different knowledge construction processes. First, whereas in the studies on peer interactions students have more chance of talking, in whole-class discussions it is expected that a greater portion of students participate only vicariously. Yet, this vicarious participation may have an effect on students' science learning. Second, in peer discussions ideas that are contrary to the ideas held by specific individuals are clearly addressed. In whole-class discussions it is not so clear. Students may perceive their ideas as addressing those of teachers, even when other peers have formulated them previously. Third, in peer interaction a discussion may be identified when opposite ideas are expressed. However, in whole-class interactions, the mere expression of opposite ideas does not configure a discussion. This is because many students' ideas normally overlap and are not properly listened or recognized by the group. Therefore, in whole class discussion opposite ideas need to be clearly identified and explored, otherwise, it is difficult to say it was a discussion. Whole-class discussions require careful teacher intervention. Consequently, whole-class discussions should be defined differently from peer discussions.

Concerning the effect of justificatory teacher-students interactions on science learning, our results are interesting. It is true that theoretically there are reasons to believe that collectively arguing for one specific idea should facilitate knowledge construction. However, results from Asterhan and Schwarz (2009) and Howe (2009) were discouraging. In the first case, authors did not find an association between consensual collective argumentation among peers and learning. The authors stated that the process of collectively arguing in favour of an idea may lead to better organisation of knowledge, but does not necessarily lead to conceptual change. Consistently, Howe (2009) does not find overall association between any type of joint construction and conceptual gains. Two points are worth noting. The first is that our study did not assess conceptual change but rather overall science learning gains. The second is that the justificatory interactions that in our study predict learning are a particular type of collaborative construction. First, they are constructions in which teachers play a central role in scaffolding students' reasoning. Second, these episodes are not characterised by consensus. They are collaborations put forward by a virtual field of possible alternative and opposite ideas, which, although not formulated clearly, are operating through teachers' interventions. This means that their effect is not accounted for by its self-explanatory potential, but by the dialectic virtual space in which these episodes are constructed. Contradiction is not always explicitly present in these episodes, but its virtual presence is suggested by teachers' interventions, creating a discursive field that pushes students'

ideas forward. In this sense, these justificatory interactions, as argumentative types of talk, are not divested of tension and conflict. It is not a consensual collaboration, explanation construction, or joint accumulative construction, but rather a process of reasoning. In this sense, they foster knowledge revision and meta-cognitive activity. Probably, these types of justificatory interactions account significantly for the effects of Mercer et al.'s exploratory talk on science learning (2004). It is worth noting that, although we called these episodes *justificatory interactions*, they should be better conceived as argumentative episodes involving more or less implicit rebuttals and counter-arguments. In any case, our results suggest that whole class teacher-students argumentative interactions have benefits for learning, even when explicit debate or discussion is not the case.

Further evidence is required about the relation between whole-class argumentation and science learning in order to determine whether this may be stated conclusively. As we mentioned, our results should be taken carefully, insofar as the sample size was too small for running the analyses that are needed to draw strong conclusions. However, it is worth noting that delayed learning gains' scores (and their distribution across groups) are striking similar to those obtained by Howe et al. (2014) using epiSTEMe materials on Forces. In any case, our results only represent a modest contribution to understanding the processes that account for the effect of certain kinds of dialogue in science learning, even when they do not clearly include explicit discussion of contradictory point of views.

## References

- Andriessen, J. E. B. & Schwarz, B. B. (2009). Argumentative design. In N. Muller Mirza & A. N. Perret-Clermont (Eds.), *Argumentation and education* (pp. 145-174). New York, NY: Springer.
- Asterhan, C. S. C. & Schwarz, B. B. (2007). The effects of monological and dialogical argumentation on concept learning in evolutionary theory. *Journal of Educational Psychology*, 99, 626-639. doi:10.1037/0022-0663.99.3.626
- Asterhan, C. S. C. & Schwarz, B. B. (2009). Argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialogue. *Cognitive Science*, 33, 374-400. doi:10.1111/j.1551-6709.2009.01017.x
- Bell, B. A., Morgan, G. B., Schoeneberger, J. A., Loudermilk, B. L., Kromrey, J. D., & Ferron, J. M. (2010, April). *Dancing the sample size limbo with mixed models: How low can you go?* Paper presented at the SAS Global Forum 2010, Seattle, WA, United States (Paper 197-2010). Retrieved from <http://support.sas.com/resources/papers/proceedings10/197-2010.pdf>
- Billig, M. (1987). *Arguing and thinking: A rhetorical approach to social psychology*. Cambridge, United Kingdom: Cambridge University Press.
- Che, C. -H. & She, H. -C. (2012). The impact of recurrent on-line synchronous scientific argumentation on students' argumentation and conceptual change. *Journal of Educational Technology & Society*, 15(1), 197-210.
- de Vries, E., Lund, K., & Baker, M. (2002). Computer-mediated epistemic dialogue: Explanation and argumentation as vehicles for understanding scientific notions. *The Journal of the Learning Sciences*, 11, 63-103. doi:10.1207/S15327809JLS1101\_3
- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84, 287-312. doi:10.1002/(SICI)1098-237X(200005)84:3<287::AID-SCE1>3.0.CO;2-A
- Duschl, R. A. & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education*, 38, 39-72. doi:10.1080/03057260208560187
- Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.
- Howe, C. (2009). Collaborative group work in middle childhood: Joint construction, unresolved contradiction and the growth of knowledge. *Human Development*, 52, 215-239. doi:10.1159/000215072
- Howe, C. (2010). *Peer groups and children's development*. Oxford, United Kingdom: Wiley-Blackwell.
- Howe, C. & Abedin, M. (2013). Classroom dialogue: A systematic review across four decades of research. *Cambridge Journal of Education*, 43, 325-356. doi:10.1080/0305764X.2013.786024
- Howe, C. & Tolmie, A. (2003). Group work in primary school science: Discussion, consensus and guidance from experts. *International Journal of Educational Research*, 39, 51-72. doi:10.1016/S0883-0355(03)00073-9
- Howe, C., Tolmie, A., & Rodgers, C. (1990). Physics in the primary school: Peer interaction and the understanding of floating and sinking. *European Journal of Psychology of Education*, 5, 459-475. doi:10.1007/BF03173132
- Howe, C., Tolmie, A., & Rodgers, C. (1992). The acquisition of conceptual knowledge in science by primary school children: Group interaction and the understanding of motion down an inclined plane. *British Journal of Developmental Psychology*, 10, 113-130. doi:10.1111/j.2044-835X.1992.tb00566.x
- Howe, C., Tolmie, A., Thurston, A., Topping, K., Christie, D., Livingston K. ... Donaldson, C. (2007). Group work in elementary science: Towards organisational principles for supporting pupil learning. *Learning and Instruction*, 17, 549-563. doi:10.1016/j.learninstruc.2007.09.004
- Howe, C., Luthman, S., Ilie, S., Guardia, P., Hofmann, R., Mercer, N. ... Taber, K. (2014). *Principled improvement in science: Forces and proportional relations in early secondary-school teaching*. Manuscript submitted for publication.
- Jansen, R. G., Wiertz, L. F., Meyer, E. S., & Noldus, L. P. J. J. (2003). Reliability analysis of observational data: Problems, solutions, and software implementation. *Behavior Research Methods*, 35, 391-399. doi:10.3758/BF03195516
- Jiménez-Aleixandre, M. P. & Erduran, S. (2008). Argumentation in science education: An overview. In S. Erduran & M. P. Jiménez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 3-28). Dordrecht, Netherlands: Springer.
- Larrain, A., Freire, P., & Howe, C. (2014). Science teaching and argumentation: One-sided versus dialectical argumentation in Chilean middle school science lessons. *International Journal of Science Education*, 36, 1017-1036. doi:10.1080/09500693.2013.832005

- Larrain, A., Freire, P., & Olivos, T. (2014). Habilidades de argumentación escrita: una propuesta de medición para estudiantes de quinto básico. *Psicoperspectivas*, 13(1), 94-107.
- Larrain, A., Howe, C., & Freire, P. (2014). *The effect of classroom oral argumentation on middle-school science learning*. Manuscript submitted for publication.
- Leitão, S. (2000). The potential of argument in knowledge building. *Human Development*, 43, 332-360. doi:10.1159/000022697
- Leitão, S. (2009). Arguing and learning. In C. Lightfoot & M. C. D. P. Lyra (Eds.), *Challenges and strategies for studying human development in cultural contexts* (pp. 221-251). Roma, Italy: Firera & Liuzzo.
- Maas, C. J. M. & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, 1, 86-92. doi:10.1027/1614-2241.1.3.86
- Mercer, N., Dawes, L., Wegerif, R., & Sams, C. (2004). Reasoning as a scientist: Ways of helping children to use language to learn science. *British Educational Research Journal*, 30, 359-377. doi:10.1080/01411920410001689689
- Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, 41, 994-1020. doi:10.1002/tea.20035
- Osborne, J. F. & Patterson, A. (2011). Scientific argument and explanation: A necessary distinction? *Science Education*, 95, 627-638. doi:10.1002/scs.20438
- Perelman, C. & Olbrechts-Tyteca, L. (1958/1969). *The new rhetoric: A treatise on argumentation* (J. Wilkinson & P. Weaver, Trans.; original title: *Traité de l'argumentation : La nouvelle rhétorique*). Paris, France: Presses Universitaires de France.
- Preiss, D., Larrain, A., & Valenzuela, S. (2011). Discurso y pensamiento en el aula matemática chilena. *Psykhé*, 20(2), 131-146. doi:10.4067/S0718-22282011000200011
- Ruthven, K., Hofmann, R., Howe, C., Luthman, S., Mercer, N., & Taber, K. (2011). The epiSTEMe pedagogical approach: Essentials, rationales and challenges. *Proceedings of the British Society for Research Into Learning Mathematics*, 31(3), 131-136.
- Scherbaum, C. A. & Ferreter, J. M. (2009). Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods*, 12, 347-367. doi:10.1177/1094428107308906
- Schwarz, B. (2009). Argumentation and learning. In N. Muller Mirza & A. N. Perret-Clermont (Eds.), *Argumentation and education* (pp. 91-126). New York, NY: Springer.
- Snijders, T. (2005). Power and sample size in multilevel linear models. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1570-1573). Chichester, United Kingdom: John Wiley & Sons.
- Sun, L., Bradley, K., & Akers, K. (2012). A multilevel modelling approach to investigating factors impacting science achievement for secondary school students: PISA Hong Kong sample. *International Journal of Science Education*, 34, 2107-2125. doi:10.1080/09500693.2012.708063
- Toulmin, S. E. (1958). *The uses of argument*. Cambridge, United Kingdom: University Press.
- van Eemeren, F. H. & Grootendorst, R. (1992). *Argumentation, communication, and fallacies: A pragma-dialectical approach*. Hillsdale, NJ: Lawrence Erlbaum.
- Venville, G. J. & Dawson, V. M. (2010). The impact of a classroom intervention on grade 10 students' argumentation skills, informal reasoning, and conceptual understanding of science. *Journal of Research in Science Teaching*, 47, 952-977. doi:10.1002/tea.20358
- Wilson, C. D., Taylor, J. A., Kowalski, S. M., & Carlson, J. (2010). The relative effects and equity of inquiry-based and commonplace science teaching on students' knowledge, reasoning, and argumentation. *Journal of Research in Science Teaching*, 47, 276-301. doi:10.1002/tea.20329
- Zohar, A. & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*, 39, 35-62. doi:10.1002/tea.10008

Fecha de recepción: Marzo de 2014.

Fecha de aceptación: Septiembre de 2014.