



ELSEVIER

Contents lists available at [ScienceDirect](#)

International Journal of Educational Research

journal homepage: www.elsevier.com/locate/ijedures



Argument relevance and structure. Assessing and developing students' uses of evidence

Fabrizio Macagno

ArgLab, IFL, Faculdade de Ciências Sociais e Humanas, Universidade Nova de Lisboa, Av. de Berna 26C, 1069-061 Lisboa, Portugal

ARTICLE INFO

Article history:

Received 18 January 2016

Accepted 1 July 2016

Available online xxx

Keywords:

Argumentation

Relevance

Argument structure

Evidence use

Dialogue

ABSTRACT

The purpose of this paper is to show whether the two crucial dimensions used for assessing the quality of argumentation, argument-as-a-product (argument structure) and argument-as-a-process (relevance), are interrelated, and how they can be used to assess the effect of argumentative mode on students' arguments. To this purpose, a twofold coding scheme will be developed, aimed at capturing: a) the argumentative function of evidence use and b) the dialogical relevance of evidence use. A study will be described in which students' use of evidence is elicited in two distinct argumentative modes (dialogical vs. non-dialogical). According to the results, in the dialogical mode students tended to use evidence in a more sophisticated way from both argument evaluation perspectives.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

The role of argumentation in students' interactions and learning is increasingly becoming a central component of education (Erduran, Ozdem, & Park, 2015; Muller-Mirza & Perret-Clermont, 2009). One of the most important goals of education is to provide people with the capacities to assess the available information, select the most relevant and adequate evidence, and make judgments and decisions based thereon. These aims constitute the core of argumentation theory, a field of studies that addresses the analysis of natural arguments in discourse (Walton, 2006). The theoretical instruments developed in the cross-disciplinary field of argumentation theory can be of fundamental importance for enhancing the students' critical thinking skills and improving classroom interactions (Carey, 2000; Muller-Mirza & Perret-Clermont, 2009; Osborne, Erduran, & Simon, 2004).

This paper focuses on a specific dimension of argumentation that is essentially related to the understanding and development of students' knowledge, namely the use of evidence. Epistemological understanding is essentially related to argument (Kuhn, Cheney, & Weinstock, 2000; Kuhn & Udell, 2007). The use of data for argumentative purposes (i.e. to support, assess, question, or refute a claim) reveals and affects student's epistemological background (Sandoval & Millwood, 2005). The way students use data and evidence shows how they have interpreted and evaluated such pieces of information, and how the latter have become or become part of their background knowledge. The interpretation, assessment, and justification of a piece of evidence, or rather the critical stand towards it, is mirrored and encouraged by how an individual coordinates it with claims and other concurring or conflicting evidence (Kuhn, 1993, 2002). The dialectical assessment of a piece of information, shown by its use in constructing arguments or questioning or attacking arguments and claims, is related to its understanding (Nussbaum & Edwards, 2011). Failure to connect evidence to a claim suggests failure to interpret

E-mail address: fabriziomacagno@hotmail.com (F. Macagno).

<http://dx.doi.org/10.1016/j.ijer.2016.07.002>

0883-0355/© 2016 Elsevier Ltd. All rights reserved.

the former, distinguish it from a point of view, and address it critically (Jimenez-Aleixandre, Rodriguez, & Duschl, 2000; McNeill & Krajcik, 2008; McNeill, Lizotte, Krajcik, & Marx, 2006; Sandoval & Millwood, 2005). Improper use of evidence risks leaving students with beliefs that they are not able to justify and assess (Osborne, Erduran, Simon, & Monk, 2001).

The development of more sophisticated uses of evidence is essentially connected with the strategies for developing argumentation skills (Kuhn, 1993, 2002). Argumentative, collaborative dialogues have been shown to be an extremely effective teaching strategy (Alexopoulou & Driver, 1996; Baker, 1999; Driver, Newton, & Osborne, 2000; Kuhn, 1992; Light & Glachan, 1985; Murray, 1982; Nussbaum, 2008a, 2008b) both for improving students' critical skills (Koballa, 1992; Osborne, 2010) and addressing background knowledge (Hewson, 1992; Limón, 2001). In dialogues, students are confronted with alternative views, and thus encouraged to take a critical stand towards the evidence they are presented with, in order to analyze the reasons to prefer a point of view or a piece of evidence over another (Osborne et al., 2001). On this argumentative perspective, education should construct students' knowledge through argumentative interactions (Baker, 1999). By addressing opposite views, countering claims, and rebutting conflicting evidence, students have been shown to acquire not only the knowledge of scientific concepts, but the grounds on which such concepts are based (McNeill & Krajcik, 2008; Osborne et al., 2001; Sandoval & Millwood, 2005).

The development of argumentation skills, and more precisely of the argumentative uses of evidence, depends on the assessment of an argument, namely determining what counts as a better or more sophisticated argument or component thereof. However, argument can mean both the product, namely the quasi-logical support (justification) of a claim (Erduran, Simon, & Osborne, 2004; Means & Voss, 1996; Zohar & Nemet, 2002), and the process, or rather the (dialogical or individual) argumentative debate. In education, most of the empirical research has focused on the product (Kuhn & Udell, 2007), neglecting the concept of goal of an argument, namely for what discursive purpose it has been advanced (Kuhn, Shaw, & Felton, 1997; Kuhn & Udell, 2003, 2007). However, arguments cannot be considered solely as quasi-logical connections between premises and conclusion (Tindale, 1999). They are reasons provided to support or challenge a viewpoint that is subject to doubt, using premises that the interlocutor is presumed to accept (Walton, 2006). The structural completeness, logical soundness, and strength of an argument is only *one* evaluative dimension. The other dimension is about the purpose of an argument, namely how the speaker's arguments are related to the interlocutor's ones or address the topic of the discussion, and how an argument can contribute to moving forward the dialogue (Walton, 2004a, 2008). In argumentation theory, this fundamental aspect of argument evaluation is analyzed under the label of "relevance" (Dascal, 1979; Giora, 1997; Grice, 1975; Tindale, 1999; Walton & Macagno, 2007; Walton, 2004b).

The purpose of this paper is to analyze the relationship between argument product and process, namely between its quasi-logical structure and its relevance (or discursive/pragmatic goal), in two distinct argumentative modes (dialogical vs. non-dialogical). The focus will be on the argumentative quality of evidence use, namely on how students use evidence for argumentative purposes. To this purpose, a twofold coding scheme, aimed at capturing both aspects of argument quality, will be used to assess the quality of students' evidence use in both modes. The results of the two distinct argument-quality coding criteria in the two argumentative conditions will be compared.

2. Literature review

In order to assess and develop the best strategies for improving students' argumentation skills, it is of fundamental importance to establish what counts as a more sophisticated argument (more specifically in this case, a more sophisticated use of evidence). Evidence is data – consisting in measurements, the authority of experts or qualified sources, or experiments – that can be used to support a point of view (Aikenhead, 2005; McNeill, 2010). To be used as the grounds for a claim, evidence needs to be interpreted so that the relevant argumentative and cause-effect relations are brought to light. Epistemological understanding can be at the same time be mirrored by and developed through students' arguments (Kuhn et al., 2000; Kuhn & Udell, 2007).

The evaluation of the uses of evidence is essentially connected with the assessment of argument quality. In order to establish what counts as a better use of a component of an argument, it is necessary to rely on a theory of argument quality. In the education literature, evidence has been assessed mostly based on its structural function, namely its "logical" function of supporting or justifying a claim. Maloney and Simon pointed out how the development of children's argumentation depends on their ability to evaluate evidence in terms of its adequacy, its relevance, and its source (Maloney & Simon, 2006; p. 1823). These three dimensions concern the force of evidence (source) and its relationship with the claim (relevance and adequacy), which is ultimately analyzed the Toulmin's Argument Pattern (henceforth TAP) (Erduran et al., 2004; Toulmin, 1958). Similarly, in (Sandoval & Millwood, 2005), the quality of evidence use was assessed according to three correlated criteria: 1) sufficiency; 2) rhetorical reference; and 3) conceptual quality (see also McNeill, 2010). All the three criteria assess the same aspect of an argument, namely the connection between justification and claim. The first criterion, addressing the structure of arguments, is used to judge whether students cited enough of the relevant data to justify their claims. Rhetorical reference concerns how evidence is referred to in students' explanations. The last criterion, conceptual quality, is used to assess whether or not evidence reasonably supports a claim (Sandoval & Millwood, 2005). Also in this case, the structural dimension is the one that is taken into account for determining the quality of evidence use. Other approaches assess the quality of the use of a premise or evidence referring to the notion of "relevance to the claim" (Schwarz, Yair, Julia, & Merav, 2003; Zembal-Saul, Munford, Crawford, Friedrichsen, & Land, 2002). However, this notion is often not defined and is measured only through the TAP model.

The assessment of students' arguments has been focused mostly on the logical (structural) dimension of an *argument*, while the investigation of the process of *argumentation* (discursive and pragmatic dimension) has mostly emphasized the type of and the relationship between the various discourse moves (units) (see for instance Reznitskaya et al., 2001). In particular, reference to the interlocutor's point of view has been used as a measure of the quality of argumentation (Kuhn, 1991). On this perspective, the purpose of an argument (process) is not only to support one's view, but more importantly to address the interlocutor's one, by challenging, weakening, or accepting, taking into account, or referring to it (Felton & Kuhn, 2001; Kuhn, Goh, Iordanou, & Shaenfield, 2008; Kuhn & Udell, 2003; Walton, 2006, 2008). The interlocutor's position needs to be tackled in order for the argument to have a dialogical function, namely modifying the acceptability of a claim (Felton & Kuhn, 2001; Felton, 2004; Kuhn & Udell, 2003; Kuhn et al., 2008).

Even though the argument structure is assessed and measured independent of its pragmatic purpose, the importance of the relationships between argument (product) and argumentation (process) (Johnson, 2000; Kuhn, 1991; O'Keefe, 1977) has been hinted at in several studies. Kuhn and Udell (2003) described a threefold classification system of argument quality. Arguments were distinguished in nonjustificatory (with no argumentative force in the discussion), nonfunctional (focusing on tangential aspects of the problem) and functional arguments (addressing core aspects of the problem). This proposal points out the crucial importance of the topic of the discussion in determining the purpose (and consequently the quality) of the argument. However, the idea of functionality or justificatory force seems to be related to the ultimate claim and not to the interlocutors' moves. The purpose is defined structurally and a priori, and not dialogically. A different approach to this issue is provided by Clark and Sampson, who integrated the structural analysis of an argument (using the TAP), the dialogical dimension of argument-as-a-process, and an analysis of the quality of the grounds (Clark, Sampson, Weinberger, & Erkens, 2007; Clark & Sampson, 2008). Thus the quality of evidence use was assessed considering how it was coordinated with other evidence to pursue distinct dialogical and argumentative purposes (Sampson & Clark, 2008). However, in this case, the dialogical dimension was used to assess the opposition of the points of view (namely how the interlocutor's view was opposed and challenged) in order to be modeled within the TAP model.

The aforementioned studies seem to acknowledge the importance of integrating process and product, but two problems arise relative to measuring the argumentation (pragmatic) dimension of an argument and its structure. The first problem concerns the argumentative (process) notion of purpose of an argument. Argumentation can be measured and analyzed in terms of the relationship between the interlocutors' moves. Since the purpose of an argument is to increase the acceptability of a viewpoint and weakening conflicting positions, the relationship between an argument and the specific issue under discussion needs to be accounted for. Both in dialogical and non-dialogical settings, the general topic of the discussion is pursued through specific sub-issues. In dialogues, the interlocutors need to take into account the interlocutor's moves and arguments, and build thereon to move the discussion forward (Walton, 2004a, 2004b). However, this relation is not measured nor analyzed in educational studies. In education, the fundamental concept of *relevance* of an argument is often mentioned but neither clearly defined nor measured (Kelly & Takao, 2002) or connected with the evaluation of arguments (Reznitskaya et al., 2001).

The second problem concern the argument structure and the purpose of the elements (data and evidence) used to support a claim. The aforementioned studies mainly rely on the TAP, but do not investigate how a piece of evidence can be used within an argument and why certain uses are better than other ones. Moreover, since the structure of an argument (product) is often investigated independent of its dialogical or pragmatic purpose (process), it is hard to analyze the different quality of evidence used to attack another's position.

3. Research questions and theoretical background

This paper addresses two interrelated questions: 1) Are the argument-as-a-product and argument-as-a-process dimensions related in the assessment of students' argumentation? 2) How does the dialogical mode affect the quality of evidence use? The two research questions are crucially interdependent, as only after establishing what counts as a better use of a piece of evidence from both perspectives it is possible to design or test methods for helping students integrate evidence in their arguments (Mayweg-Paus, Macagno, & Kuhn, 2015). The first research question is functional to the second one, namely evaluating the best argumentative mode that can be implemented for encouraging more sophisticated evidence uses. The issue of evidence use evaluation can be addressed by taking into account the advances in the area of argumentation theory, in which a twofold criterion is used: the structural function and the pragmatic purpose (namely its *relevance*).

3.1. Structural uses of evidence

The structure of an argument has been broadly analyzed in education using the Toulmin's argument structure (Erduran et al., 2004; Toulmin, 1958). According to Toulmin, an argument can be represented as an interconnected set of 1) a claim (C) (which will be referred to also as conclusion or position), 2) data (D), 3) a warrant connecting claim and data ("since W"), 4) backings (B) substantiating the warrants. The structure of Toulmin's model can be represented in Fig. 1.

This reading of Toulmin's scheme is aimed at distinguishing two distinct argumentative uses of evidence. The Data is considered as the premise that is directly linked to the conclusion, and provides the factual and relevant reason to draw the conclusion (the Claim). For example, if the Claim is "Cigarettes should be banned," a possible Data is "Cigarettes damage the lungs." The Backing is here conceived as a second-level premise, a support that can be given to an element of an argument

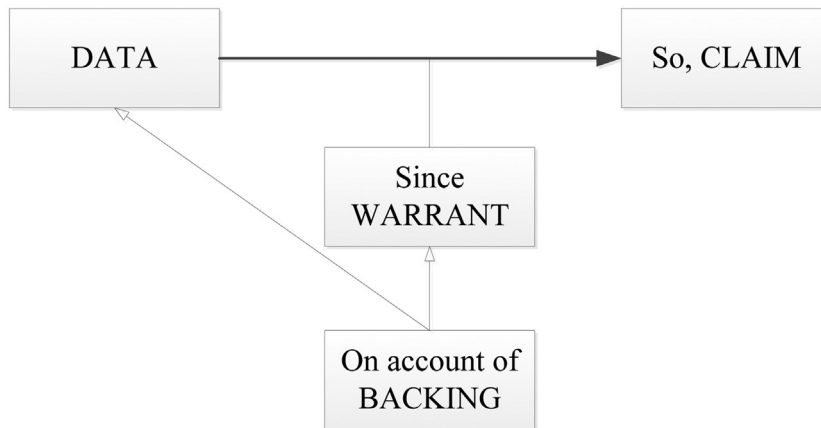


Fig. 1. Toulmin's structure of argument analysis.

(the Data or the Warrant) that constitutes the reason directly grounding the claim. For example, to support the aforementioned Data, the piece of evidence “Dr. Smith, expert in the field of respiratory diseases, claims that smoking causes lung cancer.” Evidence, on this perspective, can be used argumentatively for two distinct purposes, namely supporting a claim (positive use) or rejecting an argument (negative use).

Evidence can be used for supporting a Claim in two ways, either as a Backing to support the Data or the Warrant, or a Data to ground a claim directly. In the first case, the relationship between evidence and conclusion is indirect. Evidence needs to be interpreted, and the arguer needs to elaborate and advance the premise that a specific piece of evidence can support. On the contrary, in the second case evidence is directly related to the conclusion. However, in this case, no interpretation thereof is given. Scientific data are used without being elaborated and connected to the conclusion. For this reason, the most sophisticated use of evidence consists in its use as a backing, as it presupposes an interpretation (Sandoval & Millwood, 2005).

Evidence can be also used for attacking different components of an argument, i.e. the claim, the data, the warrant, or the backing. Evidence can be used as data or backing supporting a contrary conclusion (namely supporting a counter-claim). Otherwise, it is aimed at attacking the grounds of the argument (underminer), namely either the interpretation of the evidence, or the evidence itself (Pollock, 1970, 1974; Walton, 2006). As maintained in previous studies aimed at evaluating the quality of students' argumentation in distinct argumentative modes (Macagno, Mayweg-Paus, & Kuhn, 2015; Mayweg-Paus et al., 2015), distinct argument structures show different levels of argumentative quality. In particular, evidence used to support directly the claims (including contrary claims) reveals a lower argumentative quality (evidence is used without further interpretation), while the use of such pieces of information as a backing or as an underminer mirrors a more sophisticated quality use.

3.2. Relevance of evidence

Given the essential pragmatic nature of argumentation (Felton & Kuhn, 2001; Walton, 2006; Weinstock, Neuman, & Glassner, 2006), the purpose of the use of an argument or evidence within an argumentative discourse.¹ becomes of crucial importance. In argumentation, this relationship is investigated as a relevance relation. An argument (or more generally a discourse move) is generally considered as relevant to a goal-directed conversation to the extent that it is functional to the goal of such a discourse or part of discourse (Walton, 2004b). However, it is necessary to distinguish two aspects of argumentative relevance.

Relevance (see Blakemore, 2002; Chapter 5) can be analyzed in terms of local and global relevance (Macagno, 2008). Local relevance represents the *coherence* of a speech act (which we can refer to also as a dialogue move) within a sequence of moves (Dascal, 1979; Giora, 1997). A move is locally relevant when its content is connected with the content of the previous and following ones uttered by the speaker. For example, “*nicotine has been shown to relieve anxiety*” can be used relevantly when it is intended to support the claim that “*smoking can have beneficial effects*.” The same piece of evidence would be less relevant or not relevant at all if used to ground a different claim (for instance concerning the economic consequences of smoking). A piece of evidence is thus locally relevant when it instantiates a logical pattern of reasoning leading to the intended claim (Walton, 2003).

¹ More precisely, in this paper the notion of “discourse” will refer to both non-dialogical (when a speaker is not confronted with a physical opponent, engaging in individual argumentation) and dialogical (the speaker engages in a discussion with a physical opponent) discourse.

Global relevance concerns the purpose of the discourse. In a dialogical context, it refers to the intention that is revealed through the moves of the interlocutors, and concerns the reference to what the interlocutor intends to communicate. In a non-dialogical context, it refers to the relationship between the move and the general claim that the speaker intends to support, or the topic that he intends to address. In this sense, global relevance takes into account how the single move contributes to support the more general communicative intention of the speaker (*informativeness*). For example, a speaker can use the aforementioned move “*nicotine has been shown to relieve anxiety*” relevantly within a context in which the effects of smoking on health are discussed. However, the same move would be less related or no connected at all to previous moves addressing a different topic (such as the respiratory diseases caused by smoking or cases of smokers who died young). In such cases, this piece of evidence would hardly contribute to the informativeness of the text, reducing possible uncertainties (Giora, 1988, 1997). The speaker fails to interpret it and to indicate how it contributes to the goal of the conversation. The result is that in order to make the move relevant (coherent and informative), the interlocutor needs to reconstruct the various missing links between the dialogical context and the move itself, resulting in very high processing efforts (Wilson & Sperber, 2004).

The relevance of a piece of evidence can be defined according to distinct standards. In a dialogue, evidence can be used in a wrong way, i.e. to support the contrary conclusion, showing the speaker's wrong interpretation of it. Evidence can be irrelevant for two reasons. It can be not functional to supporting the claim or the argument or provides too little weight to it (uninformative). Otherwise, it can be not connected to the previous moves, failing to comply with the well-formedness of the discourse (incoherent) (Giora, 1985, 1997). The irrelevant uses of evidence show that the speaker has understood (to different extents) that the piece of evidence can support one of the two conflicting viewpoints, but fails (to varying degrees) to interpret it and connect it coherently with the rest of the discourse.

4. Method

In order to address the two aforementioned research questions, two coding schemes have been developed. One is intended to capture the argument-structural function of a piece of evidence; the other detects how a piece of evidence is used in a dialogue to move it forward. These two distinct criteria are intended to assess the quality of evidence use. The second step of this paper is to address the second research question using the proposed coding schemes. To this purpose, a study has been designed aimed at analyzing how the argumentative dialogical mode affects the quality of evidence use by adult students.

To assess the quality of students' uses of evidence in distinct argumentative modes, we compared adolescents' arguments concerning a social issue – *banning cigarettes in the United States* – in a peer-peer argumentative dialogue and in an individual essay. Students were assigned different positions on this issue (pro and cons), which they were requested to support using arguments and evidence. All participants received a set of cards with short pieces of evidence, which they were instructed to use at their will. The number of pieces of evidence supporting a position, the opposing one, and a neutral view on the topic were equal (Table 1).

The impact of the dialogical setting on the critical evaluation of evidence – measured through the quality of the use of the pieces of evidence – was tested in the dialogical and the non-dialogical mode. To this purpose, we developed a twofold coding system, aimed at: 1) capturing the function of a piece of evidence within the structure of an argument; and 2) measuring from an argument-as-a-process point of view whether the students used the pieces of evidence in an argumentatively relevant way. The results of analysis of the argument-as-a-product dimension – discussed in a separate paper (Mayweg-Paus & Macagno, 2016) – are compared with the one relative to the pragmatic, argument-as-a-process dimension. The purpose is to inquire into the differences and the possible relations between product and process and to investigate their development through the same or similar activities grounded on a specific argumentation mode.

4.1. Participants and design

We investigated a sample of 37 students (54.1% female) from two biology classes (10th graders) at a public High school in New Jersey (hereinafter, Class A and Class B). Their mean age was $M = 15.62$ ($SD = 0.92$), from 37 participants, 48.6% were Hispanic (Latino), 27% were African-American (black), 18.9% were American (white), 2.7% were Asian, 2.7% were Indian. Classes were balanced regarding their grade point average (GPA) on a 4-point scale, $F(1,35) = 0.522$, $p = 0.40$, *ns*, with Class A $M = 2.27$, $SD = 0.61$ and Class B $M = 2.1$, $SD = 0.82$, as well as their biology performance in per cent in particular, $F(1,35) = 0.757$, $p = 0.30$, *ns*, Class A $M = 74.15\%$, $SD = 13.88\%$ and Class B $M = 77.52\%$ $SD = 8.62\%$. The experimental manipulation was realized in 1×2 -crossover design with students.

4.2. Materials and procedure

We developed an evidence list/set of evidence cards containing 15 different pieces of evidence related to the issue of banning cigarette sales, divided into three groups: pro, contra and neutral. Six pieces of evidence stemmed from a list previously developed by Khait and Kuhn (in prep.). Another four pieces were slightly modified versions of evidence naturally used by a random sample of interviewed laypeople. The remaining 5 pieces were developed based on Internet research.

Table 1
Evidence List – Classification System.

Evidence Type	Purpose	Example
Anecdotal (single case) <i>Reference to a specific case</i>	Con	Georg Harrison, a musician for the Beatles, was a smoker and died of lung cancer in the age of 58.
	Neutral	A man called John Stasser used to smoke, but he was not addicted to cigarettes. He simply enjoyed smoking.
	Pro	A woman named Helen Reichert lives in NYC; she is 108 years old and has been smoking half a pack of cigarettes every day for over 80 years.
Authority <i>Reference to a source like scientists, government, institutes</i>	Con	The government emphasizes that smoking is one of the factors enhancing the probability of getting heart attacks.
	Neutral	Dr. F.R. Moore, president of a major cancer center, claims that smoking, combined with other causes, may lead to lung cancer, but there are many other factors.
	Pro	Scientists have shown a strong genetic influence of people living long and healthy lives, whether or not they smoke.
Popularity (Popular Opinion) <i>Reference to the majority/common sense</i>	Con	A survey showed a large majority of people support banning cigarettes sales.
	Neutral	A survey showed a large majority of people think smoking is an important issue.
	Pro	A survey showed a large majority of people think anyone has a right to decide if they want to smoke.
Laboratory evidence (Authority) <i>Reference to a study by describing the design/findings</i>	Con	Several major studies indicate a relation between smoking cigarettes and lung health.
	Neutral	Laboratory research shows that people use smoking to help them cope when facing highly stressful situations.
	Pro	Studies show the nicotine in cigarettes causes fast-acting chemical reactions in your brain that relieve anxiety and nervousness.
Statistics: descriptive (induction) <i>Reference to numbers and data</i>	Con	Each year, an estimated 443,000 people die from smoking or exposure to secondhand smoke, and another 8.6 million live with a serious illness caused by smoking.
	Neutral	An estimated 17 million Americans try to quit smoking each year, and about 8% of them succeed.
	Pro	Thousands of farmers in the U.S. make their living from farming tobacco leaves, and the tobacco industry contributes an average \$16.5 billion to the economy each year

The study took place within a 90 min school lesson. In a preliminary phase, students were provided with two questions asking for their view on the topic. Their answers were used to match students to opposing-site pairs (students holding a specific view were preferentially instructed to defend this position in dialogues). Then, cards with 15 different pieces of evidence were handed out to them and they were asked to read the information on the cards carefully. The intervention was carried out in two phases. In phase 1, students in class A were instructed to write an essay outlining their view on the topic. Students in class B were randomly assigned to opposing-site pairs and were asked to discuss the topic face-to-face. In phase 2, students in class B were instructed to write the essay and, consequently, students in class A were asked to discuss the topic in opposing-site pairs. See Fig. 2 for an overview.

This design was aimed at detecting possible influences of the order of the argumentative modes on the use of evidence.

4.3. Independent and dependent variables

The study was aimed at measuring the effect of the independent variable, namely the argumentative mode (dialogical vs. non-dialogical), on the quality of students' use of evidence (dependent measure) (research question 2). The quality of students' argumentation is measured through the aforementioned two interrelated dependent criteria, the argument-structure function and the dialogical relevance of evidence. The two criteria are not interdependent; however, according to the literature, they shall provide together a measure of the argumentative quality of evidence use (Walton, 1999, 2006). The first criterion provides an evaluation of the sophistication of the argument use within an argument structure. The second criterion provides an assessment of the discourse coherence and informativeness of a move expressing a piece of evidence. According to our hypothesis, the two criteria measure distinct aspects of the same variable, and, for this reason, they should not lead to conflicting outcomes (research question 1).

As a first step to coding the dialogs and essays were divided into idea units. From the perspective of linguistic and argumentation studies, idea units (Asterhan & Schwarz, 2009; Jucks & Paus, 2012) correspond to discourse segments (Grosz & Sidner, 1986; 178) or dialogue moves (Levin & Moore, 1977; Mann, 1988), namely units of discourse expressing a communicative intention of the speaker.



Fig. 2. Crossover Design for Class A and Class B.

Idea units were further divided into on-task units, i.e., addressed to the argumentation task, or off-task units (e.g., “we are a good team”). Then the on-task units were segmented in units either referring to evidence or not with a good interrater agreement for roughly 50% of the dialogs and essays (Krippendorff’s $\alpha = 0.94$ for both tasks) (Krippendorff, 2004). The further coding took into account only to the on-task and evidence related units.

4.3.1. The argument-structure function of evidence use

The structural use of evidence was coded according to categories drawn from previously mentioned Toulmin’s model of argument. In particular, the different positive uses of evidence (supporting a position) were distinguished from their negative ones (rejecting an incompatible claim). Relative to the positive uses, the evidence units used as a Toulmin’s backing – namely a support given to a premise leading to the wanted conclusion – were coded as “Support argument.” The evidence units used to support a position directly – namely as Data instead of as a backing – were coded as “Support position.” Concerning the negative uses of evidence, the evidence units used to counter an argument were labeled “Weaken argument.” Evidence used to directly attack to the interlocutor’s evidence was name “Weaken evidence.” In this latter case, the interlocutor’s argument is undermined by countering the evidential grounds on which it stands (Mayweg-Paus et al., 2015). The evidence used to support an incompatible claim (position), namely used as a counterargument, was coded using the codes of positive uses of evidence, inasmuch as such use corresponds to supporting a position (Table 2).

4.3.2. The relevance of evidence use

For the purpose of analyzing and assessing the use of evidence, a complementary coding system was developed, aimed at capturing how the evidence units were used from a discourse-coherence perspective. This code was intended to capture

Table 2
Argument as a product (functional) scheme.

Code	Description	Example for Use in Dialog	Example for Use in Essay
Supp_Pos ($\alpha = 0.82$) (e), $\alpha = 0.79$ (d))	Evidence (in italics) is used to support directly the claim but it is not related with Data or Warrant in an argument. It is used as a premise instead of as a backing. This code also refers to a reply given to the interlocutor in which the speaker gives contrary evidence without explaining how it is related with the other’s view.	A. So listen, you have to understand, right? <i>Each year, an estimated 443,000 people die from smoking or exposure to secondhand smoke, and another 8.6 million live with a serious illness caused by smoking. What are you going to say about that?</i> (DVR_19B)	There are people who want to banned cigarettes but nobody Talk about that <i>There is a guy who smoke cigarettes & die at the age of 58. A research shows that cigarettes have a strong relationship with lung health.</i> (D24_2B)
Supp_Arg ($\alpha = 0.75$) (e), $\alpha = 0.73$ (d))	Evidence is used to back up Data (in bold), giving strength to it. It is indirectly related with the speaker’s claim, as it supports a line of reasoning.	B. I don’t think so, like smoking is not the only reason like people think that only smoking is the reason like the asthma, or like lung cancer but it is not that, <i>Dr. F.R. Moore, president of a major cancer center, said that, that lung cancer is caused by many reasons</i> (DVR_24B)	There are many reasons to why it should be banned, not only because its bad for you but it can be addicting to people , likes an <i>estimated 17 million Americans try to quit smoking each year and about 8% of the succeed.</i> Which proves that when most people smoke, they get addicted and cant stop, and then they will be harming their lungs or other factors. (D007_1A)
Weaken_Arg ($\alpha = 0.81$) (e), $\alpha = 0.82$ (d))	Evidence is used to attack the interlocutor’s argument providing a reason not to accept it.	A. People may have, they could make a decision on whether they wanna smoke or not, but smoking releases smokes into the air and it affects not only them but everybody around them. B. No because there is no one, ah, number 6 says that <i>scientists have shown a strong genetic influence of people living long and healthy lives, whether or not they smoke. It doesn’t only depend on whether they smoke or not, that’s the thing.</i> (DVR_24B)	A lot of people believe that smoking can lead to a short life a cause various sickness , but not always. <i>Scientist have shown that a strong genetic influence of people living a long healthy life.</i> (D012_2A)
Weaken_Ev ($\alpha = 0.85$) (e), $\alpha = 0.79$ (d))	Evidence is used to attack the interlocutor’s use of evidence (underlined), undermining it.	B. My point is like, smoking do bad, but is smoke that bad? Like people think is that right? Like that woman Helen Reichert in NYC; she is 108 years old and she still smokes like half a pack of cigarettes every day and she is still healthy like, like a man, like a woman like 24–25. A. Yeah, that a thing that you can’t say, <u>you may have a person that lives like until 200 smoking</u> , but you can find someone that died, like a person, he was in the Beatles, a musician for the Beatles, he dies at 58 because of smoking, so everybody’s person is not the same (DVR_24B)	Although <u>Helen Reichert is 108 and has been smoking 80 years</u> that doesn’t mean nothing. Not every one person is alike. Just like she lived to be 108, <i>Georg Harrison lived to be 58.</i> (D024_1B)

whether the student understands the evidence (detecting whether the evidence unit can actually support the speaker's conclusion), and whether he interprets it (detecting whether he connects it to the rest of the discourse or the dialogue). To this purpose, argumentative relevance was coded according to four variables, R+, Scarcely Relevant, R-, and Wrong.

"R+" is the code for evidence units that support the wanted conclusion or address the interlocutor's viewpoint or move by providing a coherent reply. Evidence units that support the position contrary to the speaker's intended conclusion were coded as "Wrong." For example, a speaker is using a piece of evidence that supports the contrary view with the intent of providing grounds to his own position. "R-" represents the code for the use of evidence that is irrelevant for supporting a position. The speaker, for example, is providing a piece of evidence that does not contribute to supporting the claim advanced, or fails to address the topic or claim discussed by the interlocutor. "Scarcely Relevant" indicates that the speaker uses a piece of evidence without addressing the context, failing to connect it with the interlocutor's move or the previous sentences. While the piece of evidence can support the intended conclusion, it is unrelated to the conversation or the context. The Wrong evidence was intended to capture the basic understanding of the evidence; the R- and Scarcely Relevant were introduced to analyze the deeper understanding and the interpretation of the evidence (the student can connect it to what he intends or is expected to support). The coding is represented in Table 3.

The relevant, irrelevant, scarcely relevant, and wrong evidence units were computed in relation to the total number of evidence used. Interrater reliability was tested for 50% of the sample and it was found satisfying for both dialogs (Krippendorff's $\alpha = 0.79$) and essays (Krippendorff's $\alpha = 0.81$).

Table 3
Categories of Argumentative Relevance.

Code	Description	Example for Use in Dialog	Example for Use in Essay
R+	Evidence (in italics) is directly linked to the claim (in bold) and supports it.	B. I am against banning because there are people that are like 100 years old, and sometimes even more, and they have smoked for 80 years and they are still alive (D007_A)	There are many reasons to why it should be banned, not only because its bad for you but it can be addicting to people , likes an estimated 17 million Americans try to quit smoking each year and about 8% of the succeed. Which proves that when most people smoke, they get addicted and cant stop, and then they will be harming their lungs or other factors. (D007_1A)
Scarcely Relevant	Evidence supports the claim of the student, but is unrelated to the interlocutor's or the speaker's previous dialogue moves.	B. Well, a woman from NYC, she is 108 years old and has been smoking half a pack of cigarettes every day for over 80 years and she is still living. What are you going to say about that? A. What am I going to say about that? I will tell you right now. Look. <i>Studies show the nicotine in cigarettes causes fast-acting chemical reactions in your brain that relieve anxiety and nervousness, right?</i> (DVR19_B) C. But listen, do you listen to music, do you know Georg Harrison, a musician for the Beatles, was a smoker and died of lung cancer in the age of 58, bro. What are you going to say about that? A. Ah, hold on. Ah. <i>Laboratory research shows that people use smoking to help them cope when facing highly stressful situations.</i> (DVR19_B)	There are people who want to banned cigarettes but nobody Talk about that There is a guy who smoke cigarettes & die at the age of 58. A research shows that <i>cigarettes have a strong relationship with lung health.</i> (D24_2B) Before I start I will like to state a fun fact: Did you know that A women by the name Helen Reichert is 108 years old and had been smoking half Packs of cigarettes every day for 80 years. (D24_2A)
R-	Evidence does not support the claim or the argument at all or provides little weight to it.	A. But how do you feel about <i>the 108 old lady who decided to smoke for 80 years</i> , so . . . (used to counter the view that "everyone has the right to decide to smoke or not") (D002_A)	Commercials on cigarettes give information on what cigarettes are made out of. (used to support the view "I would tell them that I believe cigarettes should be banned and that I am one hundred percent certain of my point of view") (D008_1A) <i>Studies show that the methane found in cigarettes are also found in cat urine and dog poop.</i> (used to support the view "I would tell them that I believe cigarettes should be banned and that I am one hundred percent certain of my point of view") (D008_1A)
Wrong	A piece of evidence supporting a different claim is used erroneously to back up the student's conclusion.	B. <i>Studies show the nicotine in cigarettes causes fast-acting chemical reactions in your brain that relieve anxiety and nervousness</i> (used to support "smoking should be banned") (D007_B)	<i>Studies also show the nicotine in cigarettes causes fast-acting chemical reactions in your brain that relieve anxiety and nervousness.</i> (used to support the view "I would tell them that I believe cigarettes should be banned and that I am one hundred percent certain of my point of view") (D008_1A)

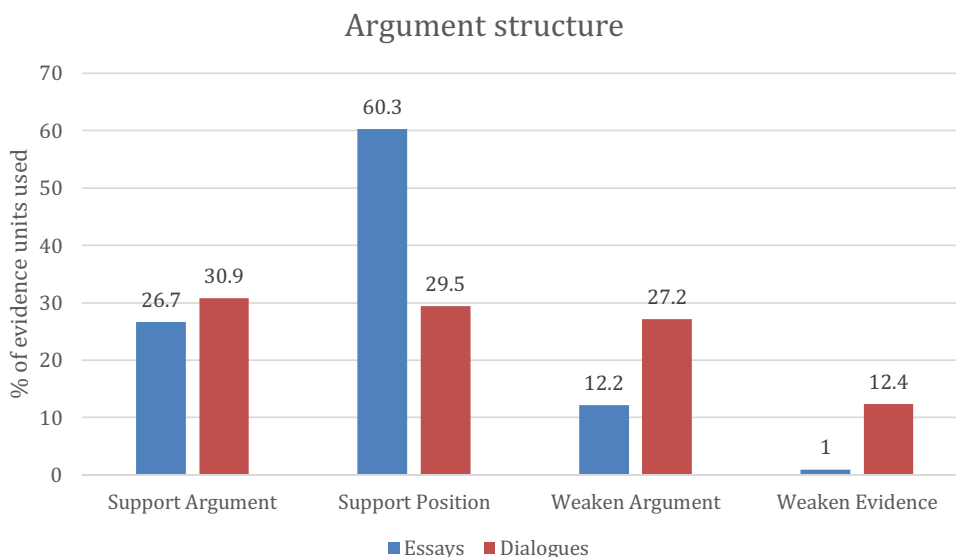


Fig. 3. Structure of evidence use according to argumentative mode (Non-dialogical and Dialogical).

5. Results

To take random effects into account and increase the generalizability of results, we employed linear mixed-effect models (Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013). To determine the potential effects of the argumentative mode and the order of intervention (as represented by an interaction of the mode) on the structural function and the relevance of evidence use, we included the argumentative mode as a fixed effect and the subjects and the dyads as random effect in the following analyses. After establishing that the order had no effect on the evidence use, the data were pooled together²

The number of on-task units differed between the two different argumentative modes, $t(34) = -4.01$, $p < 0.001$, $d = 1.05$, with dialogs being longer ($M = 20.35$, $SD = 12.30$) than essays ($M = 10.76$; $SD = 4.03$). Thereby, $M = 3.73$ ($SD = 2.28$) of the on-task units in the dialogs and $M = 2.53$ ($SD = 1.73$) of the on-task units in the essays referred to evidence. The following analyses were based on the proportions of each category of the distinct two coding schemes to the number of on task units referring to evidence. To make data suitable for further analyses, we then performed a logit-transformation of these proportions to use the transformed values.

5.1. The argumentative structure of evidence use

When engaging in the two distinct argumentative modes, students showed different uses of evidence from a functional (argument structure) perspective. We distinguish between positive uses (supporting a claim or a position) and negative uses (rejecting another's argument or position). Relative to the positive use, 26.7% of the evidence in the essays and 30.9% in the dialogs was used to support an argument, 60.3% of the evidence in the essays and 29.5% in the dialogs was used to support a position. Concerning the negative uses, 12.2% of the evidence in the essays and 27.2% in the dialogs was used to counter the other's argument. Moreover, 1.0% in the essays and 12.4% in the dialogs were used to counter the other's evidence. These results, discussed in detail in the complementary study (Mayweg-Paus & Macagno, 2016), are summarized in Fig. 3.

No significant differences were detected concerning the category "support argument." However, the analysis of the data showed a significant effect of the argumentative mode on the category "Support position" and the negative uses of

² There was no specific influence of the order on on-task units in the two different modes: ($t(71) = -1.46$, $p = 0.15$, ns). There was no specific influence of the order on the structural use of evidence. There was no specific influence of the order of intervention on the positive ($t(70) = -0.81$, $p = 0.42$, ns) or negative ($t(70) = 1.42$, $p = 0.16$, ns for counterargument; $t(70) = 1.29$, $p = 0.20$, ns for counter-evidence) uses of evidence. There was no specific influence of the order on the relevance of evidence use: $R+$: $t(33) = 1.70$, $p = 0.10$, ns (fixed coefficient: 5.09); $Wrong$: ($t(34) = -0.80$, $p = 0.43$, ns (fixed coefficient: -4.90); $Irrelevant$: $t(33) = -1.36$, $p = 0.18$, ns (fixed coefficient: -5.24).

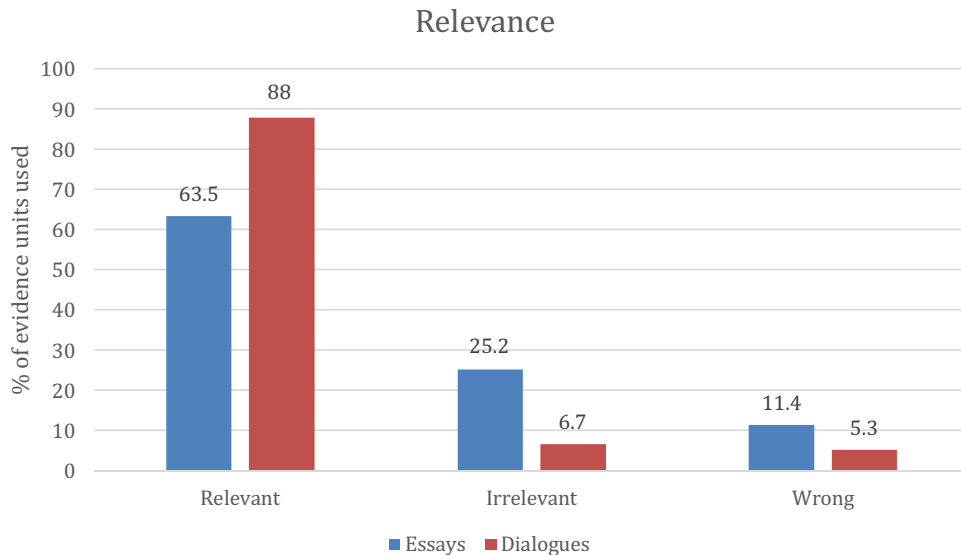


Fig. 4. Relevance of evidence use according to argumentative mode (Non-dialogical and Dialogical).

evidence.³ The analysis revealed that students tended to use more evidence to counter the opposing arguments, and more evidence to attack opposing evidence in dialogs than in essays.

5.2. The relevance of evidence use

The analysis of the relevant uses of evidence by the students (Code R+) showed that evidence was used more relevantly in the dialogs ($M = 3.36$, $SD = 2.25$) than in the essays ($M = 1.24$, $SD = 3.72$, $t(34) = -3.76$, $p < 0.001$, $d = 0.69$). The analysis of the non-relevant uses was conducted by distinguishing the cases in which evidence is used in a wrong way (Wrong) from the ones in which it is used without it being connected to a claim (R-) or the topic/context (Scarcely Relevant). According to the analysis, there were no significant differences concerning the pieces of evidence wrongly used between the two argumentative modes ($t(34) = 1.42$, $p = 0.16$, *ns*). The numbers of R- and Scarcely Relevant evidence use (incoherence due to uninformativeness or discourse incoherence) were subsumed under a common category that we referred to as “Irrelevant.” This category represents the uses of evidence that are not functional to the argumentative dialogue for different reasons. The analysis showed that students’ irrelevant evidence use differed between dialogs and essays, $t(34) = 2.96$, $p < 0.01$, $d = 0.61$. Students used evidence in a more irrelevant way in the essays ($M = 3.86$, $SD = 1.60$) than in the dialogs ($M = 2.25$, $SD = 3.37$).

The results can be summarized in the following figure (Fig. 4): in the essays, 63.5% of the evidence was coded as relevant, 25.2% as irrelevant, and 11.4% as wrongly used. In the dialogs, 88.0% of the evidence was classified as relevant, 6.7% as irrelevant, and 5.3% as wrongly used.

5.3. Quality of evidence use and argumentation mode

The aforementioned statistical data provide an overview of the distinct argumentative behavior of the students involved in the two argumentative modes. The outcomes of the structural analysis are coherent with the ones drawn from the relevance analysis, showing that in the dialogical mode students tend to use evidence in a more elaborated way, interpreting more the pieces of evidence. From an argument structure perspective, they use them to support an argument (backing) or attack a contrary view or position; from a relevance point of view, they connect them to the rest of the discourse, namely they understand how they can be used in an informative and coherent way (Giora, 1985, 1997).

The relationship and the complementarity of the two coding schemes for assessing and detecting the different argumentative quality of the two modes can be clearly shown in a qualitative analysis. The significant differences underscored by the statistical analysis can be illustrated by comparing two prototypical samples of argumentation drawn

³ In the essays students used more often evidence to support directly their position ($M = 0.84$, $SD = 3.97$) than in the dialogs ($M = -1.74$, $SD = 3.28$), $t(71) = 2.86$, $p = 0.01$, $d = 0.71$, (AIC = 406.5, BIC = 422.6, fixed coefficient: -2.99). Students used more evidence to weaken the opposing arguments in dialogs, ($M = -2.07$, $SD = 2.75$) than in the essays ($M = -3.48$, $SD = 2.74$), $t(71) = -2.66$, $p = 0.01$, $d = 0.51$, (AIC = 359.7, BIC = 375.8, fixed coefficient: -0.13). Students used more often evidence to weaken the evidence supporting the opposing position in the dialogs (undercut) ($M = -3.31$, $SD = 2.74$) than in the essays ($M = -4.49$, $SD = 0.64$, fixed coefficient: -1.14), $t(71) = -2.74$, $p = 0.01$, $d = 0.59$. These results are reported and discussed in (Mayweg-Paus & Macagno, 2016).

Table 4
Example of an essay.

Move	Contribution	Code
1	Each year, an estimated 443.000 people die from smoking or exposure to second hand smoke and another 8.6 million live a serious illness caused by smoking.	Scarcely relevant
2	If I was invited to join a discussion about banning cigarette sales in the US, I would tell them that I believe cigarettes should be banned and that I am one hundred percent certain of my point of view.	Other
3	I personally know a young boy who has/had a hole in his throat since birth. That hole was due to his mother's cigarette addiction before and while she was pregnant with him.	Scarcely relevant
4	Commercials on cigarettes give information on what cigarettes are made out of.	R-
5	Studies show that the methane found in cigarettes are also found in cat urine and dog poop.	R-
6	Studies also show the nicotine in cigarettes causes fast-acting chemical reactions in your brain that relieve anxiety and nervousness.	Wrong
7	My grandmother can smoke up to a pack of cigarettes a day.	R-
8	And reading these facts makes me even more nervous of her health in the future.	Other

from the two modes. The twofold argument-structure and discourse-coherence (relevance) analysis allows understanding how evidence was used, and why a specific use should be considered as less sophisticated from an argumentative perspective. The essay (Table 4) is strongly characterized by pieces of evidence unconnected with the claim or the context.

From a structural perspective, the various pieces of evidence (1 and moves from 3 to 7) all contribute to supporting the claim directly (support position), which can be regarded as a sign of failure to interpret the information used. The structural analysis brings to light how the student uses the evidence to support or counter a position. However, it does not show whether this support or attack is coherent with the dialogue, or is appropriate considering the content of the piece of evidence used. From the point of view of argument-as-a-process, the student fails to develop a uniform or coherent line of argument. He fails to connect the various issues that the pieces of evidence mentioned address, employing them as if they were topically coherent. Instead of interpreting them and using them, he provides a list of unconnected and unelaborated possible supports.

The prototypical dialogue shows a different attitude towards the use of evidence from an argument-as-a-process perspective (Table 5).

In this dialogue, the interlocutors tend to connect the pieces of evidence to a topic. They interpret the evidence, and connect it to a premise (such as at 7, first piece of evidence used) or to the interlocutor's argument (such as at 7, second piece of evidence used, or 9). Students tend to take into account the conflicting point of view, showing how the evidence they are using rebuts the opposing position (such as 7, 8, and 12). More importantly, they take into consideration in their arguments the interlocutor's argument and use of evidence (at 10; 15). They address not only the general topic, but also the specific claim and the specific premises used by the interlocutor, connecting their moves (in this case the specific use of a piece of

Table 5
Example of a dialog.

Turn	Dialog Partner	Contribution	Code
1	A	Ok. Do you wanna ban cigarettes or not?	Other
2	B	Yeah, I wanna ban them	Other
3	A	Why?	Other
4	B	Because they kill people and get them hearth cancer, I mean lung cancer,	Other
5	A	Ok, what else?	Other
6	B	They kill people and what else . . . cause billions . . . it says economy . . . something. Do you think they should ban cigarettes? They should right?	Other
7	A	Yeah. I don't think we should ban cigarettes. Because it helps people. Like my sister, she smokes so when she is feeling stressed, she just smoke and she get cooler. I see that way but . . . I don't think they should ban it, it helps people. Ad there was a fact here it says that smoking cigarettes isn't bad because you can live long and healthy life.	R+ R+
8	B	Hmm I am	Other
9	A	And there was one about this lady who is 100 years old and she has been smoking cigarettes for 80 years and she still alive.	R+
10	B	This musician from the Beatles, Georg Harrison, was a smoker and died of lung cancer in the age of 58.	R+
11	A	But all depends on the smoker. If you are addicted to you obviously it is bad but if you know how can control it, it can be good.	Other
12	B	Hold on. I have another one. An estimated 17 million Americans try to quit smoking each year, and about 8% of them succeed. It's not the only one. Each year, an estimated 443,000 people die from smoking	R+
13	A	It's scary here	Other
14	B	or exposure to secondhand smoke, and another 8.6 million live with a serious illness caused by smoking.	R+
15	A	Ok, John, he smokes, he used to smoke, but he was not addicted to cigarettes. He simply enjoyed smoking. So it is not bad, it not that bad	R+
16	B	People die.	Other
17	A	Smoking cigarettes help people cope with highly stressful situations, which is all right I am done for myself	R+

evidence) not only to a general goal of the discourse, but more specifically to the conclusion and the interpretation of the evidence advanced by the other party.

The qualitative analysis of the prototypical essay and dialogue show how both coding schemes can be used to capture distinct and complementary aspects of students' argumentation. A lower elaboration and understanding of evidence is mirrored by both a less sophisticated structural use of the information given and in a lower connection thereof to a topic or claim to support. While in dialogues the students tend to connect evidence to the interlocutor's moves or their own claim to defend, essays often show a pattern of unconnected remarks, only sometimes addressing a common topic.

8. Discussion

By comparing the quality of students' uses of evidence in the dialogical and non-dialogical mode, we have found that evidence was used in a more sophisticated way in dialogues under both an argument-as-a-product and an argument-as-a-process perspective. This finding is coherent with previous works in education, which underscored that argumentative discourse activity enhances the development of argument skills (Felton & Kuhn, 2001; Kuhn & Moore, 2015; Kuhn & Udell, 2003; Kuhn et al., 2008, 1997; Reznitskaya et al., 2001).

The quality of student's argumentation was assessed by taking into account the use of evidence for both supporting an uncertain conclusion and contributing to the goal of the conversation. In education, these two dimensions have been always considered separately. The "structural" assessment of an argument was investigated considering the completeness of the TAP (Erduran et al., 2004; Sampson & Clark, 2008). The dialogical dimension was analyzed in terms of addressing conflicting or opposing points of view (Felton & Kuhn, 2001; Kuhn et al., 1997). However, for evaluating argumentation quality and more specifically the quality of use of evidence, both views provide essential criteria, which have been developed and combined herein.

The structural criterion (product) has been developed by integrating the dual goal of argumentation within the evaluation criteria. Building on (Felton & Kuhn, 2001; Kuhn & Udell, 2003; Macagno et al., 2015), the argumentative quality of a piece of evidence was structurally assessed based on its dual argumentative function, namely backing an argument or supporting a claim, and weakening an argument or a contrary backing (evidence). Structurally, quality has been defined based on the understanding of opposing views and the interpretation of the evidence. The dialogical dimension (process), complementary structural one, was developed building on the insights of (Kuhn & Udell, 2003). The criterion of relevance (intended as informativeness or better argumentative coherence) was introduced to distinguish whether a move is connected to its claim and the other moves of the discourse, namely whether it contributes to the purpose of a discourse or a dialogue.

According to the initial hypothesis and theoretical assumption, the two dimensions of argumentative quality are interrelated. Sophisticated uses of evidence presuppose an interpretation not only of the evidence itself, but also of the claim that it is intended to support or the argument that is aimed at weakening. On the contrary, evidence used in a less sophisticated way is also less interpreted (or not interpreted at all), leaving up to the interlocutor the burden of finding or reconstructing the relationship with the discourse or the dialogue (if any). In this sense, interpretation is mirrored by both structure and dialogical coherence and informativeness.

The results show that the two coding schemes are coherent in their results. The quality of argumentation in the two modes has been assessed independently using the two distinct criteria (coding schemes). The results are consistent, pointing out that students tend to use evidence in a structurally more sophisticated way and in a more relevant fashion in the dialogical mode. They have shown to connect better evidence to premises or opposing arguments, and to address not only the general issue at stake, but also the specific interpretation of the evidence advanced in the previous moves.

From a structural perspective, we have found a crucial difference in the kind of argumentative use of evidence (used for both purposes of supporting or attacking a viewpoint) (Mayweg-Paus et al., 2015; Mayweg-Paus & Macagno, 2016). Relative to the positive uses of evidence, the dialogical and the non-dialogical argumentative modes showed a significant difference. In comparison with the dialogues, in the essays students tended to support their positions using more often evidence to directly support the conclusion, as if it were the explicit premise of an argument. This type of evidence use does not presuppose any elaboration and interpretation of the information provided, and can be considered as a less elaborated argumentative strategy, as more subject to possible criticisms and misunderstandings (Macagno & Walton, 2014, Chapter 6 and 7). Also relative to the negative uses of evidence, the students showed a different behavior in the two modes. In the dialogic setting students tended to use more often than in the essays evidence as an underminer. They focused their attacks more on the arguments and the backings available to or advanced by the interlocutor. In this sense, they tended more to address the evidence and the interpretation thereof advanced by the other party, engaging more deeply and authentically in a critical analysis of the positions and the evidence at their disposal.

From an argument-as-a-process point of view, the findings show that in essays students tend to use evidence in a less discourse-coherent fashion. In their essays, students tended more than in dialogues to fail to connect the various pieces of evidence used to the context. On the contrary, in dialogues students tended more to interpret the evidence and use it to address the interlocutor's claims and arguments, or to pursue a uniform line of argument (addressing a common topic). This finding is coherent with the outcome of the analysis of the structural dimension of evidence use. In order to be used relevantly, a piece of evidence needs to be interpreted and connected to the context (namely the previous sentences or the previous interlocutor's moves).

The argument-as-a-product and the argument-as-a-process perspectives can thus provide coherent and complementary views on the same phenomenon. The mere structural perspective cannot capture whether the students used the evidence to achieve a specific goal, namely whether the use of such pieces of evidence actually contributed to the purpose of the discourse or dialogue. However, the structural scheme can show why and how a move was relevant or irrelevant considering its place within the structure of an argument.

8.1. Limitations and implications

The limitation of the relevance coding consists in the fact that it depends on the content analysis. Relevance cannot be determined independently of analyzing the content, as it is necessary to establish what discursive or dialogical goal the speaker is pursuing, and how he is doing it. The analyst is thus involved also in interpretative activities, namely: 1) interpreting what the student intends to claim; 2) assessing whether the student's interpretation of the evidence is correct; and 3) determining what is the topic or the claim discussed in the previous moves.

These findings and results suggest the crucial importance of developing curricula based on a dialogic approach (Kuhn, Hemberger, & Khait, 2014) for developing students' argumentative skills (Kuhn & Crowell, 2011), in particular the ones used for distinguishing, evaluating, assessing, and using different types of evidence. The dialogical setting influences positively the development of argument skills, considered both from the point of view of argument as a product and argument as a process.

Acknowledgements

Fabrizio Macagno would like to thank the Fundação para a Ciência e a Tecnologia for the research grants IF/00945/2013, PTDC/MHC-FIL/0521/2014, and PTDC/IVC-HFC/1817/2014.

References

- Aikenhead, G. S. (2005). Science-based occupations and the science curriculum: Concepts of evidence. *Science Education*, 89(2), 242–275.
- Alexopoulou, E., & Driver, R. (1996). Small-group discussion in physics: Peer interaction modes in pairs and fours. *Journal of Research in Science Teaching*, 33(10), 1099–1114. [http://doi.org/10.1002/\(SICI\)1098-2736\(199612\)33:10<1099:AID-TEA4>3.0.CO;2-N](http://doi.org/10.1002/(SICI)1098-2736(199612)33:10<1099:AID-TEA4>3.0.CO;2-N).
- Asterhan, C., & Schwarz, B. (2009). Argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialog. *Cognitive Science*, 33(3), 374–400. <http://dx.doi.org/10.1111/j.1551-6709.2009.01017.x>.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <http://dx.doi.org/10.1016/j.jml.2007.12.005>.
- Baker, M. (1999). Argumentation and constructive interaction. In P. Coirier, & J. Andriessen (Eds.), *Studies in writing: Vol. 5. Foundations of argumentative text processing* (Vol. 5, pp. 179–202). Amsterdam: University of Amsterdam Press.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <http://dx.doi.org/10.1016/j.jml.2012.11.001>.
- Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Cambridge: Cambridge University Press.
- Carey, S. (2000). Science education as conceptual change. *Journal of Applied Developmental Psychology*, 21(1), 13–19. [http://dx.doi.org/10.1016/S0193-3973\(99\)00046-5](http://dx.doi.org/10.1016/S0193-3973(99)00046-5).
- Clark, D., & Sampson, V. (2008). Assessing dialogic argumentation in online environments to relate structure, grounds, and conceptual quality. *Journal of Research in Science Teaching*, 45(3), 293–321.
- Clark, D., Sampson, V., Weinberger, A., & Erkens, G. (2007). Analytic frameworks for assessing dialogic argumentation in online learning environments. *Educational Psychology Review*, 19(3), 343–374. <http://dx.doi.org/10.1007/s10648-007-9050-7>.
- Dascal, M. (1979). Conversational relevance. In A. Margalit (Ed.), *Meaning and use* (pp. 153–174). Dordrecht: Reidel.
- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84(3), 287–312. [http://dx.doi.org/10.1002/\(SICI\)1098-237X\(200005\)84:3<287:AID-SCE1>3.3.CO;2-1](http://dx.doi.org/10.1002/(SICI)1098-237X(200005)84:3<287:AID-SCE1>3.3.CO;2-1).
- Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's Argument Pattern for studying science discourse. *Science Education*, 88(6), 915–933. <http://dx.doi.org/10.1002/sc.20012>.
- Erduran, S., Ozdem, Y., & Park, J.-Y. (2015). Research trends on argumentation in science education: a journal content analysis from 1998 to 2014. *International Journal of STEM Education*, 2(1), 1–12.
- Felton, M., & Kuhn, D. (2001). The development of argumentative discourse skill. *Discourse Processes*, 32(2), 135–153. http://dx.doi.org/10.1207/S15326950DP3202&3_03.
- Felton, M. (2004). The development of discourse strategies in adolescent argumentation. *Cognitive Development*, 19(1), 35–52. <http://dx.doi.org/10.1016/j.cogdev.2003.09.001>.
- Giora, R. (1985). Notes towards a theory of text coherence. *Poetics Today*, 6(4), 699–715.
- Giora, R. (1988). On the informativeness requirement. *Journal of Pragmatics*, 12(5–6), 547–565.
- Giora, R. (1997). Discourse coherence and theory of relevance: Stumbling blocks in search of a unified theory. *Journal of Pragmatics*, 27(1), 17–34.
- Grice, P. (1975). Logic and conversation. In P. Cole, & J. Morgan (Eds.), *Syntax and semantics 3: speech acts* (pp. 41–58). New York: Academic Press.
- Grosz, B., & Sidner, C. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3), 175–204.
- Hewson, P. W. (1992). Conceptual change in science teaching and teacher education. *Research and curriculum development in science teaching*. Madrid: National Center for Educational Research, Documentation, and Assessment, Ministry for Education and Science.
- Jimenez-Aleixandre, M. P., Rodriguez, A. B., & Duschl, R. (2000). Doing the lesson or doing science: Argument in high school genetics. *Science Education*, 84, 757–792. [http://dx.doi.org/10.1002/1098-237X\(200011\)84:6<757:AID-SCE5>3.0.CO;2-F](http://dx.doi.org/10.1002/1098-237X(200011)84:6<757:AID-SCE5>3.0.CO;2-F).
- Johnson, R. (2000). *Manifest rationality*. Hillsdale: Lawrence Erlbaum Associates.
- Jucks, R., & Paus, E. (2012). Different words for the same concept: Learning collaboratively from multiple documents. *Cognition and Instruction*, 31(2), 227. <http://dx.doi.org/10.1080/07370008.2013.769993>.
- Kelly, G. J., & Takao, A. (2002). Epistemic levels in argument: An analysis of university oceanography students' use of evidence in writing. *Science Education*, 86(3), 314–342. <http://dx.doi.org/10.1002/sc.10024>.
- Koballa, T. R. (1992). Persuasion and attitude change in science education. *Journal of Research in Science Teaching*, 29(1), 63–80. <http://dx.doi.org/10.1002/tea.3660290107>. <http://www3.interscience.wiley.com/cgi-bin/fulltext?ID=112752996&PLACEBO=IE.pdf&mode=pdf> [Retrieved from].

- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Thousand Oaks: Sage. <http://dx.doi.org/10.2307/2288384>.
- Kuhn, D., & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents' thinking. *Psychological Science*, 22(4), 545–552.
- Kuhn, D., & Moore, W. (2015). Argumentation as core curriculum. *Learning: Research and Practice*, 1(1), 66–78. <http://dx.doi.org/10.1080/23735082.2015.994254>.
- Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development*, 74(5), 1245–1260.
- Kuhn, D., & Udell, W. (2007). Coordinating own and other perspectives in argument. *Thinking & Reasoning*, 13(2), 90–104.
- Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentative reasoning. *Cognition and Instruction*, 15(3), 287–315.
- Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development*, 15(3), 309–328.
- Kuhn, D., Goh, W., Iordanou, K., & Shaenfield, D. (2008). Arguing on the computer: A microgenetic study of developing argument skills in a computer-supported environment. *Child Development*, 79(5), 1310–1328.
- Kuhn, D., Hemberger, L., & Khait, V. (2014). *Argue with me: Argument as a path to developing students' thinking and writing*. New York: Wessex Press.
- Kuhn, D. (1991). *The skills of argument*. New York: Cambridge University Press.
- Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review* 62(2), 155–178. . [Retrieved from] <http://search.proquest.com/docview/61286294?accountid=11077>.
- Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education*, 77(3), 319–337. <http://dx.doi.org/10.1002/sce.3730770306>.
- Kuhn, D. (2002). What is scientific thinking and how does it develop? In U. Goswami (Ed.), *The wiley-Blackwell handbook of childhood cognitive development, second edition* (pp. 497–523). Oxford: Blackwell Publishers Ltd.. <http://dx.doi.org/10.1002/9781444325485.ch19>.
- Levin, J. A., & Moore, J. A. (1977). Dialogue-games: Metacommunication structures for natural language interaction. *Cognitive Science*, 1(4), 395–420. [http://dx.doi.org/10.1016/S0364-0213\(77\)80016-5](http://dx.doi.org/10.1016/S0364-0213(77)80016-5).
- Light, P., & Glachan, M. (1985). Facilitation of individual problem solving through peer interaction. *Educational Psychology*, 5(3–4), 217–223. <http://dx.doi.org/10.1080/0144341850050305>.
- Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and Instruction*, 11(4–5), 357–380. [http://dx.doi.org/10.1016/S0959-4752\(00\)00037-2](http://dx.doi.org/10.1016/S0959-4752(00)00037-2).
- Macagno, F., & Walton, D. (2014). *Emotive language in argumentation*. New York: Cambridge University Press.
- Macagno, F., Mayweg-Paus, E., & Kuhn, D. (2015). Argumentation theory in education studies: Coding and improving students' argumentative strategies. *Topoi*, 34, 523–537.
- Macagno, F. (2008). Dialectical relevance and dialogical context in Walton's pragmatic theory. *Informal Logic*, 28(2), 102–128.
- Maloney, J., & Simon, S. (2006). Mapping children's discussions of evidence in science to assess collaboration and argumentation. *International Journal of Science Education*, 28(15), 1817–1841.
- Mann, W. C. (1988). Dialogue games: Conventions of human interaction. *Argumentation*, 2(4), 511–532.
- Mayweg-Paus, E., & Macagno, F. (2016). How dialogic settings influence evidence use in adolescent students. *Zeitschrift Für Pädagogische Psychologie*, 30(2–3), 121–132.
- Mayweg-Paus, E., Macagno, F., & Kuhn, D. (2015). Developing argumentation strategies in electronic dialogs: Is modeling effective? *Discourse Processes* 1–18. <http://dx.doi.org/10.1080/0163853x.2015.1040323>.
- McNeill, K. L., & Krajcik, J. (2008). Inquiry and scientific explanations: Helping students use evidence and reasoning. In J. Luft, R. Bell, & J. Gess-Newsome (Eds.), *Science as inquiry in the secondary setting* (pp. 121–134). Arlington: National Science Teachers Association Press..
- McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *Journal of the Learning Sciences*, 15(2), 153–191. http://dx.doi.org/10.1207/s15327809jls1502_1.
- McNeill, K. L. (2010). Explanations, arguments and evidence in science, science class and the everyday lives of fifth grade students. *Annual Meeting of the National Association for Research in Science Teaching* (pp. 1–26)..
- Means, M., & Voss, J. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. *Cognition and Instruction*, 14(2), 139–178.
- Muller-Mirza, N., & Perret-Clermont, A.-N. (2009). *Argumentation and education: Theoretical foundations and practices*. Dordrecht: Springer.
- Murray, F. B. (1982). Teaching through social conflict. *Contemporary Educational Psychology* 7(3), 257–271. [http://doi.org/http://dx.doi.org/10.1016/0361-476X\(82\)90031-5](http://doi.org/http://dx.doi.org/10.1016/0361-476X(82)90031-5).
- Nussbaum, M., & Edwards, O. V. (2011). Critical questions and argument stratagems: A framework for enhancing and analyzing students' reasoning practices. *Journal of the Learning Sciences*, 20(3), 443–488. <http://dx.doi.org/10.1080/10508406.2011.564567>.
- Nussbaum, M. (2008a). Collaborative discourse, argumentation, and learning: Preface and literature review. *Contemporary Educational Psychology*, 33(3), 345–359. <http://dx.doi.org/10.1016/j.cedpsych.2008.06.001>.
- Nussbaum, M. (2008b). Using argumentation vee diagrams (AVDs) for promoting argument-counterargument integration in reflective writing. *Journal of Educational Psychology*, 100(3), 549–565. <http://dx.doi.org/10.1037/0022-0663.100.3.549>.
- O'Keefe, D. (1977). Two concepts of argument. *Journal of the American Forensic Society*, 13, 121–128.
- Osborne, J., Erduran, S., Simon, S., & Monk, M. (2001). Enhancing the quality of argument in school science. *School Science Review*, 82(301), 63–70.
- Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, 41(10), 994–1020. <http://dx.doi.org/10.1002/tea.20035>.
- Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, 328(5977), 463–466. <http://dx.doi.org/10.1126/science.1183944>.
- Pollock, J. (1970). The structure of epistemic justification. *American Philosophical Quarterly*, 4, 62–78.
- Pollock, J. (1974). *Knowledge and justification*. Princeton: Princeton University Press.
- Reznitskaya, A., Anderson, R., McNurlen, B., Nguyen-Jahiel, K., Archodidou, A., & Kim, S. (2001). Influence of oral discussion on written argument. *Discourse Processes*, 32(2–3), 155–175.
- Sampson, V., & Clark, D. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education*, 92(3), 447–472. <http://dx.doi.org/10.1002/sce.20276>.
- Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction*, 23(1), 23–55. http://dx.doi.org/10.1207/s1532690xci2301_2.
- Schwarz, B., Yair, N., Julia, G., & Merav, I. (2003). Construction of collective and individual knowledge in argumentative activity. *The Journal of the Learning Sciences*, 12(2), 219–256.
- Tindale, C. (1999). *Acts of arguing: A rhetorical model of argument*. Albany: State University of New York Press.
- Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- Walton, D., & Macagno, F. (2007). Types of dialogue, dialectical relevance and textual congruity. *Anthropology & Philosophy*, 8(1–2), 101–119.
- Walton, D. (1999). Profiles of dialogue for evaluating arguments from ignorance. *Argumentation*, 13(1), 53–71.
- Walton, D. (2003). Defining conditional relevance using linked arguments and argumentation schemes: A commentary on professor Callen's article, rationality and relevancy: Conditional relevancy and constrained resources. *Michigan State Law Review*, 4(4), 1305–1314.
- Walton, D. (2004a). Classification of fallacies of relevance. *Informal Logic* 24(1), 183–185. . [Retrieved from] http://ojs.uwindsor.ca/ojs/leddy/index.php/informal_logic/article/view/2133.
- Walton, D. (2004b). *Relevance in argumentation*. Amsterdam-Philadelphia: John Benjamins Publishing Company.
- Walton, D. (2006). *Fundamentals of critical argumentation*. New York: Cambridge University Press.
- Walton, D. (2008). *Informal logic: A pragmatic approach*. Cambridge University Press. . [Retrieved from] <http://www.cambridge.org/9780521713801>.

- Weinstock, M. P., Neuman, Y., & Glassner, A. (2006). Identification of informal reasoning fallacies as a function of epistemological level, grade level, and cognitive ability. *Journal of Educational Psychology, 98*(2), 327.
- Wilson, D., & Sperber, D. (2004). Relevance theory. In L. Horn, & G. Ward (Eds.), *Handbook of pragmatics* (pp. 607–632). Oxford: Blackwell. <http://dx.doi.org/10.1016/j.pragma.2009.09.021>.
- Zemal-Saul, C., Munford, D., Crawford, B., Friedrichsen, P., & Land, S. (2002). Scaffolding preservice science teachers' evidence-based arguments during an investigation of natural selection. *Research in Science Education, 32*(4), 437–463.
- Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching, 39*(1), 35–62.