THE DIMENSIONS OF ARGUMENTATIVE TEXTS AND THEIR ASSESSMENT

FABRIZIO MACAGNO,
CHRYSI RAPANTA

Abstract

The definition and the assessment of the quality of argumentative texts has become an increasingly crucial issue in education, classroom discourse, and argumentation theory. The different methods developed and used in the literature are all characterized by specific perspectives that fail to capture the complexity of the subject matter, which remains ill-defined and not systematically investigated. This paper addresses this problem by building on the four main dimensions of argument quality resulting from the definition of argument and the literature in classroom discourse: dialogicity, accountability, relevance, and textuality (DART). We use and develop the insights from the literature in education and argumentation by integrating the frameworks that capture both the textual and the argumentative nature of argumentative texts. This theoretical background will be used to propose a method for translating the DART dimensions into specific and clear proxies and evaluation criteria.

Introduction

Educational policy documents around the world increasingly emphasize the need to develop argument literacy skills in the classroom (Newell, Beach, Smith, & VanDerHeide, 2011; Reznitskaya & Wilkinson, 2015), specifically the transversal and complex ability to support one's viewpoint and address the viewpoints of others by providing arguments, counter-arguing, and refuting counter-arguments both in oral and written forms and in various disciplines (Graff, 2003, pp. 3–4). This ability implies skills at various levels, including cognitive, metacognitive, and epistemological levels (Rapanta, Garcia-Mila, & Gilabert, 2013), which justify the effects of its development on learning outcomes (Kuhn, 2010; Osborne, 2010; Von Aufschnaiter, Erduran, Osborne, & Simon, 2008). Due to the fundamental social, educational, and critical role of argument literacy, many studies in educational research – including the whole field of argumentation and education – have focused on the analysis and assessment of various types of argumentative discourse, specifically the discourse emerging within argumentation contexts.

One crucial challenge is the assessment of the quality of an argumentative text. The concept of argumentative text does not correspond to the concept of argument itself (Azar, 1999), even though argumentative texts (and discourses) essentially involve the use of arguments. An argumentative text can be considered of low quality even if the arguments used therein are good, as such arguments may be not related to overall goal of the essay or to each other (Choi, 1988; Paglieri, 2015; Witte & Faigley, 1981). Establishing clear criteria for assessment is fundamental, as only by determining the characteristics that define a good argumentative text is it possible to design effective strategies for improving the quality of students' argumentation (Chinn, 2006; Osborne, Erduran, & Simon, 2004).

A systematic method for pursuing this goal is currently missing in the literature. This methodological gap may be due to a lack of an integration between the characteristics that make a text argumentative to provide a unique assessment of its quality and a lack of dialogue between argumentation theory and educational studies. In education and, in particular, within the area of argumentation and education, the quality of arguments has been distinguished methodologically from the quality of argumentation (Rapanta & Macagno, 2016). The negative result is that a plethora of coding schemes, rubrics, and frameworks have been proposed for analyzing students' oral and written argumentation, each capturing a different dimension thereof and thus leading to different evaluations. The development of distinct assessment criteria mirrors an underlying disagreement on what counts as an argumentative text, or rather what *aspect* thereof can provide an indication of

its quality. This selective approach has the advantage of providing precise and homogeneous measurements. However, the problem is what such measurements can be taken to indicate. The number of components of an argument or the frequence of anticipations of the viewpoints of others (or criticisms in an essay) can be precisely determined; however, the relationship between these indicators and the overall quality of an argumentative text is extremely defeasible due to its partiality.

This paper addresses the challenge of developing a method for evaluating the quality of argumentative texts that integrates different dimensions and criteria. To this purpose, a theoretical framework is proposed, based on the tools developed in the fields of argumentation theory and education and combining the models that underlie the current coding schemes. After reviewing the most important theoretical frameworks used by the existing methods in education and discourse or text analysis, we bring to light the crucial dimensions that define an argumentative text, and that thus can be used and integrated for capturing the argument quality of student contributions.

Approaches to the analysis of argumentative texts

The analysis of the quality of an argumentative text has been addressed in education, linguistics (and text analysis), and argumentation theory. In these disciplines, it is possible distinguish two distinct approaches to the evaluation of argumentative texts: the first emphasizes the *argument* dimension of an argumentative text, while the second emphasizes the *textual* features. This distinction is mirrored by an equivalent one traced in argumentation theory between the quality of an argument as a *product* and the quality of the *process* of argumentation (O'Keefe, 1977), which led to independent developments (Govier, 1992; van Eemeren & Grootendorst, 2004). As Johnson pointed out, this divide has marked two distinct subfields in argumentation, where informal logic focused on the task of developing criteria for evaluating and criticizing *arguments*, and where dialogue logic concerned the definition (and often formalization) of the rights and duties in a rational *dialogue* (Johnson, 2000, p. 291).

The first approach, concerned with the product of argumentation, i.e. arguments, was pursued in education by two distinct research trends (Cavagnetto, 2010; Clark & Sampson, 2008; Manz, 2016; Rapanta et al., 2013; Sampson & Clark, 2006). On the one hand, coding schemes have been developed aimed at capturing the *structural* completeness of an argument. Using this view, an argument is evaluated based on the presence of its components, commonly detected considering Toulmin's framework. Thus, a good argument would show a complete argument structure consisting of the

claim, the premises, the backing, the qualifier, and the counterclaims (Erduran, Simon, & Osborne, 2004; Jimenez-Aleixandre, Rodriguez, & Duschl, 2000; G. Kelly, Druker, & Chen, 1998; Osborne et al., 2004), allowing different degrees of completeness. On the other hand, this purely structural framework was integrated by taking into account other features, namely the quality of the reasons used (distinguishing, *inter alia*, among abstract, personal, or vague reasons) (Schwarz, Yair, Julia, & Merav, 2003), the relevance of the justification (disregarding the distinction between the components thereof) (Zohar & Nemet, 2002), and the epistemological quality of an argument. This latter property was measured by considering the distinct epistemic levels of the premises (ranging from general knowledge to appeals to specific theories) (Kelly & Takao, 2002; Manz, 2016), or the combination of causal relations with the relevant data (Sandoval & Millwood, 2005).

The second approach has emphasized the *procedural* view of argumentation, specifically how different standpoints are taken into consideration, challenged, weakened, and integrated in one's argument (see for instance, Felton & Kuhn, 2001; Kuhn & Udell, 2003). Using this perspective, argumentative texts are viewed as dialogue moves (Felton, 2004) manifesting crucial argument skills, particularly argument construction, justification, counterargument construction, and refutation (Rapanta et al., 2013, p. 510). The quality of an argumentative text is thus dependent on two fundamental variables: the "transactivity," or connection with the purpose of the argument (such as addressing a specific viewpoint or claim) (Felton & Kuhn, 2001), and the acknowledgment of contrary or different perspectives, assessed and weakened through a critical analysis thereof, in so-called "integrative arguments" (Nussbaum, 2003; Nussbaum & Edwards, 2011; Nussbaum & Schraw, 2007).

The structural and procedural views of argumentation bring to light some fundamental features characterizing argumentative texts. The first and most important feature is the presence of arguments; however, the evaluation of the quality of an argument needs to take into account structural, epistemic, and pragmatic criteria. As the reviewed approaches point out, a good argument is not only a logical structure that includes the essential components thereof (premises and conclusion), but also a complex textual and more importantly pragmatic unit (van Eemeren & Grootendorst, 1984) that needs to be connected with the context, the interlocutor, a communicative goal, and a cognitive construct, in which the acceptability and validity of the conclusion and the premises need to be established considering their content and thus their epistemic dimension (captured by the backings and the evidence on which the premises are based).

These aspects, emerging from the literature in education and argumentation, can be found partially in the linguistic studies that tackled the problem of

evaluating *texts* that happen to be argumentative. The studies in this field emphasize evaluation criteria complementary to the previous ones. The textuality dimension is commonly captured by indicators such as connectives (Akiguet & Piolat, 1996); the variety and completeness of rhetorical structures are considered as measures for structural quality (Azar, 1999; Stab & Gurevych, 2014), while specific linguistic marks (such as the presence of oppositive connectives or the expression of degrees of certainty or endorsement) are used to establish epistemic negotiation (reflected in the aforementioned pragmatic and epistemic dimensions) (Golder & Coirier, 1994).

Theoretical frameworks for assessing the quality of argumentative texts

The structural, the textual (or more broadly pragmatic), and the epistemic approaches have defined the quality of an argumentative text based on specific perspectives on argument and argumentation. The approaches are not exclusive; rather, they are complementary, defining categories of properties that together characterize an argumentative text, which we will refer to as "levels." Some theories tend to include into one approach (for instance, in an approach focusing on the structural level) features of argumentative texts belonging to a different level (such as the epistemic level). However, this multi-level nature of an argumentative text has never been systematically addressed, even though the need to consider its different levels of analysis has been clearly acknowledged in the literature (Akiguet & Piolat, 1996; Azar, 1999; Coirier & Golder, 1993; Golder & Coirier, 1994; Sampson & Clark, 2006, 2008; Stab & Gurevych, 2014).

The crucial challenge of this paper is to provide an approach that at the same time integrates and operationalizes these different levels. The first fundamental matter is to detect the *aspects* of such levels that can be operationalized as indicators of the quality of an argumentative text. To this purpose, we consider the most important theoretical models underlying the operationalization of the distinct levels that emerge from the education literature – and are mirrored by the literature in linguistics as textual clues. While many coding schemes have been developed, the underlying theories of what counts as an argumentative text can usually be traced to three basic frameworks (for a review, see Rapanta & Macagno, 2016; Rapanta, Garcia-Mila, & Gilabert, 2013; Nussbaum, 2003, 2011): Toulmin's argument pattern, Kuhn's dialogical approach, and the argumentation schemes theory. In the following sections, we present these three theoretical frameworks and point out their limitations, which can be overcome by integrating them.

Toulmin's argument pattern

The grounds of the structural approach to the analysis of argumentative texts used in education (see for instance Erduran et al., 2004; Osborne et al., 2004) and linguistics (see for instance Golder & Coirier, 1994) can be found in Toulmin's pioneering work entitled *The Uses of Argument* (1958). Although the book is still considered a masterpiece in philosophy for a number of reasons, it only became known in education because of the simple schematization of an argument and its main elements (see Figure 1 for the broadest interpretation of this model). This schematization or pattern is composed of six main elements distinguished based on their function within the argument. The function of the data (D) is to support a claim (C), the function of the qualifier (Q) is to moderate the epistemic validity of the claim, the function of the warrant (W) is to guarantee the logical relation between the data and the claim, the function of the backing (B) is to give support to the warrant (as per Toulmin's example) and the data (in a broader interpretation, see Mayweg-Paus & Macagno, 2016), and the function of the rebuttal (R) is to recognize conditions of exception or restrictions to the warrant. According to Toulmin (1958), all types of grounds of the claim, including data, backing, and warrants, are field-dependent in substantial arguments. Therefore, it is expected that educational researchers applying TAP for assessing arguments produced in specific contexts take into consideration the disciplinary field and its effect on the epistemic function of each TAP element, especially when it regards the grounds or type of evidence used (Figure 1).
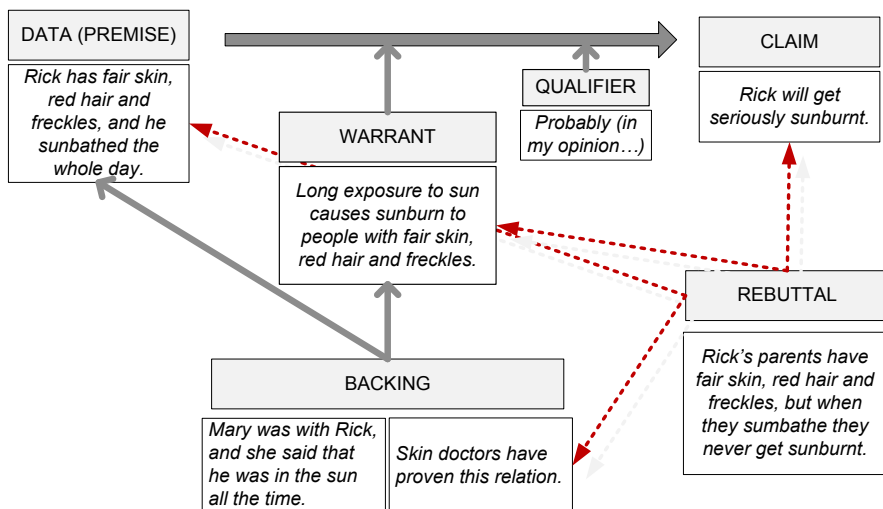

Figure 1
The TAP structure

Because of its form-focusing nature, this framework has been mainly used for assessing students' written arguments (Macagno & Konstantinidou, 2013; Sampson & Clark, 2008), but with various concerns and limitations (Rapanta et al., 2013; Sampson & Clark, 2008). Its application to classroom oral discourse is described by Erduran et al. (2004), but it has never been extensively used for this latter purpose.

Kuhn's dialogical approach

Kuhn's methodology for assessing the quality of students' arguments is based on two principles: openness to the other's argument and the use of evidence. The first principle can be described as a strategy for describing and measuring the *dialogicity* of students' arguments (an aspect that is part of the pragmatic dimension of a text), which Kuhn expresses as follows (Kuhn, 2010, p. 816):

> The first and most crucial development we look for is an increase in students' ability and willingness to attend critically to the other's argument. Until this happens, no genuine argumentation has occurred. Once they begin to listen to what the opponents have to say, the second challenge becomes constructing a counterargument that successfully weakens the force of the other's argument.

Under this view, the quality of a written argument is established based on the consideration of the other's viewpoint. Thus, one-sided arguments, or arguments aimed only at supporting one's viewpoint, are distinguished from dual-perspective arguments, in which the interlocutor's position or argument is addressed by attacking it. Finally, arguments expressing an integrative perspective are considered the most sophisticated, as they provide a balanced opinion on one's own position, including its drawbacks or the positive aspects of the opposing view (Kuhn & Crowell, 2011).

The other principle of argument quality that Kuhn develops is the use of evidence (Kuhn, 2010, pp. 816–817). The ability to integrate evidence in one's argument, or the use of evidence to attack the partner's argument, is considered as a sign of higher argument quality. The crucial difference between the use of uninterpreted evidence (a claim is true or false because the evidence says so) and the interpretation of the evidence to support or attack a position was developed by distinguishing between the following codes, presented in Table 1 (Mayweg-Paus & Macagno, 2016).

Table 1
Levels of argument quality – Use of evidence

Argument quality	Category	Description
First-order evidence	Support Position	Evidence is used to support directly a generic viewpoint (either one's own position or the viewpoint contrary to the partner's), but it is not related to an argument.
Second-order evidence	Support Argument	Evidence is used to back up an argument, strengthening it. It is indirectly related to the student's position, as it supports a line of reasoning.
Second-order evidence	Weaken Argument	Evidence is used to weaken the opposing argument by providing a reason not to accept it.
Second-order evidence	Weaken Evidence	Evidence is used to weaken directly the evidence that supports the opposing view.

This framework combines a pragmatic (dialogical) dimension capturing the *process* of argumentation with a structural and epistemic one, assessing the argument as a *product*.

Argumentation schemes theory

The last framework underlying several approaches to the assessment of the quality of argumentative texts is argumentation schemes theory (Anthony & Kim, 2015; Macagno & Konstantinidou, 2013; Metaxas, Potari, & Zachariades, 2016; Nussbaum, 2003, 2008). Walton's theory of argumentation schemes (Walton, 1995; Walton, Reed, & Macagno, 2008) is focused on the analysis and classification of Toulmin's warrants (Toulmin, Rieke, & Janik, 1984). An argumentation scheme represents the structure of the most common types of arguments in everyday conversation. They are schemes in the sense that they appear as combinations of premises leading to a conclusion based on rules of inference that capture the most common abstract justificatory relations, such as cause-effect, definition-*definiendum*, and expert opinion-acceptability. A common argumentation scheme used both in everyday and in academic situations is the *argument from expert opinion*, represented in Table 2 (Walton et al., 2008, p. 91).

Table 2
Argumentation schemes – Argument from expert opinion

Minor Premise 1	Source E is an expert in subject domain S containing proposition A.
Minor Premise 2	E asserts that proposition A (in domain S) is true (false).
Conditional Premise	If source E is an expert in a subject domain S containing proposition A, and E asserts that proposition A is true (false), then A may plausibly be taken to be true (false).
Conclusion	A may plausibly be taken to be true (false).

This scheme represents the inferential structure of the argument characterized by a specific semantic relation between premise and conclusion, namely the opinion of an expert. The premises establish the conditions for the correct use, thus distinguishing sound and acceptable arguments from the ones that cannot provide an equally strong support to the conclusion, and can be used fallaciously (Walton, 2010b). The dialogicity aspect of the scheme is captured by the set of critical questions associated to it (due to the presence of this dialogical dimension, schemes are called *argumentation*, and not simply *argument* schemes). The critical questions associated with the argument from expert opinion are the following:

1. *Expertise Question*: How knowledgeable is E as an expert source?
2. *Field Question*: Is E an expert in the field D that A is in?
3. *Opinion Question*: What did E assert that implies A?
4. *Trustworthiness Question*: Is E personally reliable as a source?
5. *Consistency Question*: Is A consistent with what other experts assert?
6. *Backup Evidence Question*: Is E's assertion based on evidence?

These questions represent the defeasibility conditions, the potentially weak points that an interlocutor can address. For example, in the scheme above, the important condition for its appropriate use is the determination of what counts as "expertise," and more precisely what is to be a reference in a domain or field. The schemes in this sense merge the structural dimension with the epistemic one: the quality of an argument is established based not only on its completeness (the presence of the components of Toulmin's pattern), but more importantly on the presence of evidence that can anticipate and preempt possible critical questions (Rapanta & Walton, 2016a, 2016b).

The schemes capture the logical and semantic quality of an argument, but also an aspect of its pragmatic level: the adequacy of the use of a scheme for pursuing a specific dialogical goal. The schemes are classified according to their pragmatic goal, distinguishing the macro-categories of purposes they can be used to support (Macagno & Walton, 2015), as shown in Figure 2.
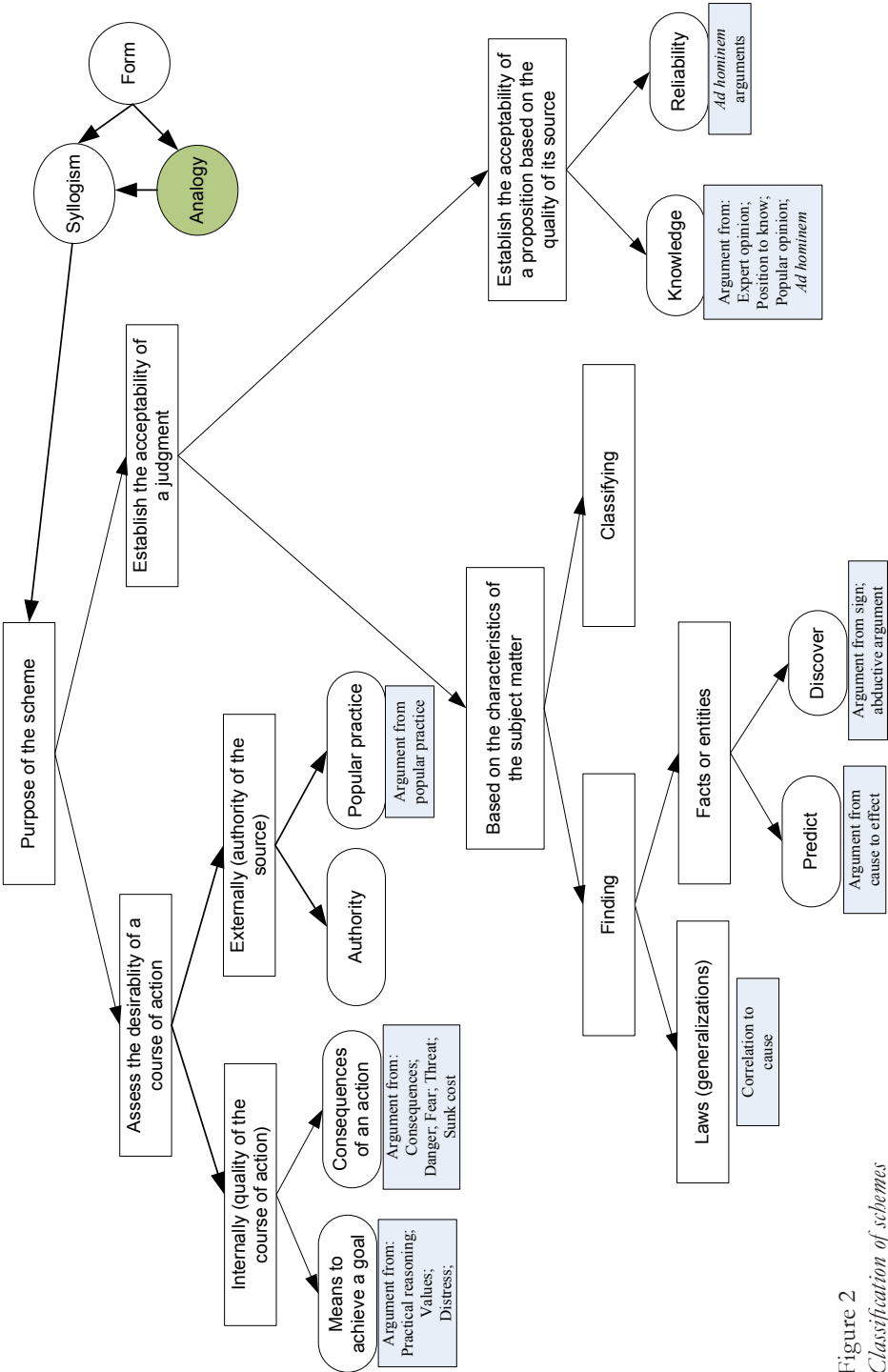
Figure 2
Classification of schemes

Argumentation schemes thus can be used for assessing when evidence use is effective for improving the soundness of an argument, going beyond the purely structural paradigm of Toulmin. The pragmatic dimension of such schemes can also be used for determining the suitability of an argument to the goal of the text.

The three theories analyzed present different frameworks and objects of analysis. Toulmin's model considers the argument as a structure; Kuhn's theory emphasizes the dialectical integration of different viewpoints; finally, the argumentation schemes theory examines the types of inference used, their soundness based on the evidence provided, and their suitability to the overarching goal. These theories provide instruments that cannot be considered as purely on one level of analysis. Toulmin's pattern is clearly structural, but the presence of backings and rebuttals involves elements belonging to epistemic and dialectical (pragmatic) levels. Kuhn's model is at the same time dialectical (pragmatic) and epistemic (related to the use of evidence). Finally, argumentation schemes are structural, but the pragmatic functions of the schemes fall into the domain of relevance (in the sense of adequacy with the purpose of the text) and the critical questions belong to both the dialogical and the epistemic levels. These theoretical frameworks are thus multi-level; however, their focus is not primarily on argumentative *texts*, but rather on arguments expressed in dialogues or writing. To integrate these frameworks and the resulting methods for operationalizing the assessment of the quality of argumentative texts, it is necessary to establish first the *aspects* of the levels identified above in the literature and the theoretical frameworks (structural, textual/pragmatic, and epistemic) that can in fact mirror it.

The dimensions of an argumentative text: the DART

To analyze the quality of an argumentative text, it is first necessary to define it. First and foremost, an argumentative text is a text, a "unit of language in use," consisting of any passage, spoken or written, that forms a unified whole as a unit of meaning (Halliday & Hasan, 1976; Witte & Faigley, 1981). This "unity," or coherence, is both at the grammatical level of (textual) cohesion and the content and pragmatic level of relevance with the overall purpose of the interaction between the speaker/writer and the hearer/reader (Giora, 1985, 1998; Macagno, 2018). An argumentative text is characterized by a goal consisting of increasing the acceptability of a doubtful viewpoint (Walton, 2006, pp. 3–4). The viewpoint can be an opinion, but also an explanation, a proposal for action, or an interest in a negotiation (Walton, 1990). The essential feature of the viewpoint in an argumentative text is its doubtfulness,

the fact that it is not accepted, or it cannot be considered as *prima facie* acceptable by the interlocutor (Walton, 1990, p. 411, 1992, pp. 185–187; see also van Eemeren & Grootendorst, 2004, p. 1). A doubtful viewpoint requires reasons for increasing its acceptability or defeating criticisms or dissents, which are expressed through arguments and counter arguments (Azar, 1999; Golder & Coirier, 1994), namely goal-directed pieces of reasoning consisting of premises supporting a conclusion through specific rules of inference.

This definition can be combined with the three levels of analysis: the structural, the epistemic, and the pragmatic (dialogical). Of these levels, the structural represents the manifestation of the other levels. When an argument is structurally complete, it is epistemically grounded through the backings or the evidence, it is logically sound because it has the premises necessary for making it a (textually) cohesive unit, and it addresses its pragmatic goal because it includes the other's viewpoints through attacks and counterarguments. However, it is already clear that the textual/pragmatic criterion described above needs to be analyzed with consideration for its various dimensions. Two dimensions – cohesion and reference to the other's viewpoint – have been mirrored in the literature at a structural level and have been operationalized through the notions of argumentation schemes or the combination of data-warrant-claim, and the concepts of counterclaim, attack, or reference to the other's view. However, as we have seen above, an argument can pursue different goals, and, as the argumentation schemes theory underscores, an argument needs to be suited to the goal that the text is intended to pursue, which is captured by the dimension of relevance.

Considering this analysis, an argumentative text can be considered under four dimensions: dialogicity, accountability, relevance, and textuality (jointly referred to as DART). Textuality mirrors the most basic requirement of the coherence of a text – the cohesion or relationship between its parts; in our case, the premises and conclusion. Relevance represents the pragmatic dimension of the coherence of a text – the relationship between an argument and the goal of the text. Dialogicity belongs to the pragmatic level, as it concerns the peculiar relationship between the interlocutors relative to the specific purpose of the argumentative text – addressing a doubt or difference. Finally, accountability consists primarily of the provision of evidence – the epistemic dimension of the notion of argument. The role and the features of each dimension will be described and justified in the following subsections.

Dialogicity/criticality

The definition of argument at the basis of an argumentative text is grounded on a dialogical dimension: the "difference" between the speakers or the "doubt" of the interlocutor, which is resolved through a specific type of

explicit and goal directed reasoning. A difference presupposes another actual or potential viewpoint, and more importantly another individual whose ideas need to be taken into account (or negotiated) for a difference to be overcome (Golder & Coirier, 1994). This dialogical aspect is commonly described in the literature in terms of "dialogicity" (Garcia-Mila & Andersen, 2007) or "criticality," through which a written argument's dialogicity is manifested (Glassner & Schwarz, 2007; Kuhn, Hemberger, & Khait, 2014; Osborne, 2010; Walton, 1989).

Critical thinking and argumentation are two terms often co-occurring in educational research studies, mainly because argumentation is thought to enhance critical thinking skills, as manifested in students' critical discourse (Osborne, 2010). According to Walton (1989), critical thinking refers to some general dispositions such as empathy and critical detachment, which are straight forwardly developed through engagement in argumentative dialogue. As Walton puts it, "the common core of basic critical thinking skills underlying critical reasoning (...) is the *key ability to look at both sides of an argument*. The structure behind this ability is the concept of argument as dialogue" (Walton 1989, p. 182, emphasis added). Criticality, which in rhetoric is often analyzed as *kairos*, or suitability to a specific audience and the views thereof (Kinneavy, 2002; Vatz, 1973), implies that the speaker takes into account the other's views, and adopts a critical stance or look at the reality, and this is an essential indicator of the quality of students' argumentative discourse.

A "critical look at the reality" means that the person accepts that reality is multiple and that different theories and values may apply to the same data and vice versa, which has also been defined as "*antilogos*" (Glassner & Schwarz, 2007). The lack of manifestation of this ability results in two of the main critical thinking flaws, defined as "my-side" bias and "makes-sense" epistemology (Perkins, Farady, & Bushey, 1991). Following Kuhn (1991), this critical thinking bias refers to the lack of ability of a speaker or a writer to accept the existence and the validity of any alternative theories to his/her own, acting based on a rather absolutist epistemological set of beliefs, which are far from the critical or evaluativist stance (Kuhn & Park, 2005). The lack of critical stance also results in the adoption of the first available view or data that "makes sense," without a rigorous analysis of its relevance, sufficiency, and acceptability. The "rhetorical" effect of the lack of criticality is the failure to persuade the audience, as the premises used cannot modify the interlocutor's attitude towards a viewpoint (Aristotle, p. 25–26).

Accountability (or use of evidence)

The second dimension of an argumentative text is related to the epistemic dimension of an argument, namely support of a potentially controversial conclusion through premises that are accepted or more acceptable (Golder

& Coirier, 1994; Sandoval & Millwood, 2005). This dimension, concerning the use of evidence (and accepted premises) to support the conclusion, has been commonly referred to as "accountability" (Erduran, Ozdem, & Park, 2015; Michaels, Connor, & Resnick, 2008; Sandoval & Millwood, 2005) or (in the literature of textual and discourse analysis) "authenticity" (Alexander, 2008; Cranton, 2001; Giddens, 1991; Long, 1996; Van Lier, 1996).

Accountability consists of a) the process of interpreting information (or phenomena) and communicating it to pursue shared understanding, and b) the possibility of assessing the reasons that speakers provide. In this sense, a text can be considered as accountable when it provides a new understanding of a state of affairs that is grounded in evidence, which makes it possible to assess its acceptability. An idea becomes an opinion if it is accountable to (i.e. understandable by and grounded in the common knowledge of) the community; an opinion becomes an argument if it is accountable to reasoning from evidence; finally, an argument becomes an acceptable argument if the evidence in which it is grounded is based on some type of shared knowledge (in the case of discipline-heavy contexts) (Michaels et al., 2008).

Accountability can be extended to include contexts in which the "sharedness" of knowledge is not always possible. Kuhn's definition of *genuine* contribution – from an argument quality point of view – broadened the concept of "accountability" to refer to the use of evidence (as opposed to pseudo-evidence). A personal theory is *genuine* when it is based on reasons that are different from both the theory itself and the reasons "contained" in the theory (Kuhn, 1991). In this sense, a student needs to interpret and use the *evidence* available to develop authentic arguments. From this perspective, accountability is related to the concept of authenticity developed in the literature of textual analysis, as the analysis, interpretation, and use of evidence leads to developing original and different arguments.

Relevance

As highlighted in the introduction of this section, an argumentative text is a text defined by its goal (to address an actual or potential difference) and its fundamental "rhetorical structure," namely the use of arguments. Both these features are pragmatic notions; they are related to the speaker's intention to pursue a communicative goal. The very concept of argument as defined above is essentially pragmatic, as it is based on the conversational goal of trying to resolve, or contend with, a difference between the interlocutors. This pragmatic dimension of an argumentative text can be captured by a crucial construct in the philosophy of language and linguistics: relevance (Reinhart, 1980).

Relevance refers to the specific quality of a move in a discourse – or more generally a unit or sequence in a text: its coherence with the joint communicative intention to which a discourse move is intended to contribute. This idea was introduced by Grice, who maintained that the participants to a conversation need to share a common communicative purpose, a common goal characterizing their verbal interaction (Grice, 1975). According to Grice, relevance (which he called "relation") is defined as appropriateness to the conversational needs (Grice, 1975, p. 47):

> Relation. I expect a partner's contribution to be appropriate to the immediate needs at each stage of the transaction. If I am mixing ingredients for a cake, I do not expect to be handed a good book or even an oven cloth (though this might be an appropriate contribution at a later stage).

Relevance is essentially related to the interpretation of the text segment and the text itself, and more specifically to the communicative goal that the interlocutors pursue (Giora, 1985, 1988; Macagno, 2018).

In an argumentative text, relevance is defined considering the goal of the dialogue, namely supporting the intended conclusion through arguments appropriate to the intended audience and to the context (Paglieri, 2015). The relevance of a unit of an argumentative text (usually labelled a "sequence") can be thus assessed considering the relationship between the argumentative function that it performs (whether as an attack, a conclusion, a counterargument, or an argument) and the dialogical context in which it is produced, namely the arguments that the speaker intends to attack, develop, specify, etc., or the other intentions to which the speaker intends to contribute (Giora, 1997; Leech, 1983; Van Dijk & Kintsch, 1983; Walton, 2004). When a sequence does or can potentially or presumably contribute to the overall goal of the text within the given context, it can be considered as relevant (Macagno, 2018, 2019). Otherwise, the sequence is evaluated as irrelevant.

Textuality

The definition of text reported above brings to light the essential dimension of the coherence of a text (see also Hanks, 1989): a text can be considered as such when the sequences constituting it form a coherent unit. Relevance captures the pragmatic dimension of this type of unity, as the sequences (and the arguments) of a text need to be aimed at a specific and unique goal that is appropriate in the given context (Macagno, 2018). Textuality can be considered as its semantic counterpart, and refers to the internal, "logical" structure of a piece of discourse that make it a coherent whole (Giora, 1997; Van Dijk, 1977a, 1977b).

Textuality is a "semantic property of discourses, based on the interpretation of each individual sentence relative to the interpretation of other sentences" (Van Dijk, 1980, p. 93). This definition is based on a logical organization of discourse topics, in which each sentence contributes to the whole as a part. However, it is purely semantic; while it can capture how the elements in an argument and the arguments themselves are organized and related to each other, it cannot capture alone the other essential dimension of an argument, namely its goal-directness (previously identified as relevance) or "point" (Schank et al., 1982; Van Dijk, 1977a). The unit of analysis of textuality is the semantic construct called a "sentence," neglecting both the communicative goal for which it is uttered and the wider cultural context and communicative context in which it occurs (Levinson, 2012, p. 107). To analyze argument quality, both relevance and textuality are needed, as they can be used to assess the complementary and fundamental aspects of an argument.

In this sense, textuality and relevance are two aspects of the same dimension of an argumentative text, namely its "coherence." The parts that are intended to form a text – and in our case an argumentative text – need to be related to a specific communicative goal (relevance) and to the other parts (textuality). The interrelation between the two dimensions of textuality and relevance can be illustrated in the examples in Table 3 below, which show how a highly pragmatically relevant piece of discourse can be poorly semantically or textually coherent, and vice versa.

Table 3
Student discourse examples to distinguish between (pragmatic) relevance and (semantic) textuality

Example 1 (+ relevance, – textuality)
(1) On the one hand, I agree with the statement (that the Mediterranean diet is recommendable) because it contains, in percentages, the correct portions. (2) On the other hand, I do not agree because the portions are extremely reduced so most people who follow it are undoubtedly hungry. (3) In short, this diet, in my opinion, should be practiced, but with more portions of everything. (4) Example: I think that eating only one or two spoonfuls of rice in one meal is very little; it should be twice that amount.
Example 2 (– relevance, + textuality)
(1) Yes, I agree with this statement, because as we know, healthy eating respects the characteristics of the individual (age, sex, health status, build, etc.) and uses meals that obey the portions indicated in the food circle. (2) The food circle corresponds to a graphic representation that helps to combine the foods that should constitute the daily meals. (3) In short, we can conclude that the Mediterranean diet presented in the food circle contributes to health promotion.

In Example 1, the student states an argument about whether the Mediterranean diet should be recommended using a strategic structure, from an argumentative point of view, known as a "balanced argument" or "two-sided argumentation" (Nussbaum & Schraw, 2007). In this sense, the student's argument is relevant to the persuasion goal of the discourse. However, from the point of view of textual coherence at a purely semantic level, the implicit or explicit connections between the sentences are poor. For example, Sentence 2 contradicts Sentence 1, as the speaker provides two contradictory opinions based on two contradictory judgments on the same state of affairs (i.e., portions in Mediterranean diet). These properties cannot be subsumed under a more generic property, nor are they related by any other semantic relation (Van Dijk, 1980). Example 2 illustrates the contrary: a contribution on the same topic that aims at being persuasive, as it forms part of a persuasive essay, but it is uncessful as the strategy used, i.e. describing the list of nutritional items composing the Mediterranean diet and the concept of the food circle, is irrelevant to the goal of the text, namely supporting a specific viewpoint (Walton, 2004). On the other hand, the main concepts used are highly interconnected with each other, which provides an acceptable discourse coherence (textuality).

The DART dimensions and the existing frameworks

The distinct dimensions of the quality of an argumentative discourse described above are summarized in Table 4.

Table 4
DART dimensional constructs that define the quality of student contributions from an argumentation point of view

Dimension	Definition
Dialogicity	The manifestation of the skill of "antilogos," namely the consideration of the other party's arguments and positions in one's own discourse
Accountability	The use of different arguments and evidence to justify (be accountable for) a viewpoint – namely an interpretation of a state of affairs.
Relevance	The coherence of an element with the purpose of the text, namely its potential effect on the overall goal of the conversation. In an argumentative setting, the effect consists of increasing the acceptability of a conclusion adequate to the context considering the interlocutors' background knowledge.
Textuality	The internal coherence of text, involving both the explicit connectedness between its units (sequences) (grammatical expression of the connections) and the semantic relations between the concepts expressed (semantic relation).

DART represents the dimensions – or variables – that can be used to evaluate the quality of student's argumentative texts. These dimensions are abstract constructs that can be hardly applicable for an objective and justifiable assessment of texts. Such abstract variables need to be operationalized, translated into a set of proxies that can be defined based on specific indicators. To this purpose, it is useful to analyze how the existing theoretical frameworks have addressed these dimensions, in order to draw on them to determine the possible indicators. Table 5 shows a summary of the degrees (partial or full) to which each one of the three proposals covers the DART dimensions.

Table 5
DART argument quality dimensions addressed by three existing frameworks

Dimensions Framework	Dialogicity	Accountability	Relevance	Textuality
Toulmin	partial	partial	–	full
Kuhn	full	partial	partial	–
Walton	partial	full	full	–

Toulmin's pattern captures the dimension of textuality, as it describes through the components of an argument the essential semantic relations or warrants that make an argument a complete textual unity. This framework also involves the epistemic dimension of accountability, captured by the concept of "backing" and its distinction from "data." However, accountability is only partially assessed, as in the TAP evidence it is monological (not accounting for the evidence used by others or against others' viewpoints). The TAP also mirrors one aspect of dialogicity, namely the presence of rebuttals. Toulmin's notion of rebuttal is much more limited than the broader one used in argumentation theory, which encompasses various types of counterarguments (Walton, 2010a). However, the concept of relevance is not addressed in this framework.

Kuhn's theory is primarily dialogical, and for this reason the textuality dimension is not mentioned. Instead, the focus of this approach is placed on the dialogicity and the accountability dimensions. As pointed out above, Kuhn distinguishes the less sophisticated argumentative strategy (i.e., "support position") from the more sophisticated ones based on both the consideration of the other's viewpoint and the critical use of evidence ("weaken argument" and "weaken evidence"). From the accountability perspective, this framework captures only partially how an argument is supported by evidence – without an account of textuality and relevance,

evidence can be used ineffectively but the argument would be still evaluated as high in its accountability dimension. The dimension of relevance is limited to reference to the other's turn, which can be useful in dialogues, but of limited use in texts.

Argumentation schemes are instruments for determining the accountability of an argument, detecting how evidence is used and how the premises are related to the conclusion. Moreover, they provide a method for assessing the relevance of an argumentative text both at the level of the elements of an argument and considering the relationship between an argument and the overall text (Macagno, 2019). The dialogicity dimension is also captured by the critical questions, even though only partially as the interlocutor is only considered in one of its aspects – as the criticizer of an argument, but not as the holder of a viewpoint.

Operationalizing the DART framework

As shown in Table 5, the three reviewed frameworks are characterized by their focus on one specific dimension of an argumentative text. In this sense, they can be used to measure one dimension in all its complexity, but the other dimensions are either only partially captured or are not considered. These models, however, can be integrated, as we will show in the following subsections.

Integrating the frameworks: Dialogicity

One of the crucial aspects of dialogicity is the integration of the other's perspective in one's own argument. This feature, however, is not captured by argumentation schemes theory or by the TAP model (Leitão, 2000): the former considers the interlocutor's view and position in terms of potential doubts, while the latter in terms of rebuttal. This gap can be addressed by considering Kuhn's approach, which is the most focused on the dialogicity dimensions and offers a range of possible dialogic elements (Table 5).

According to Kuhn (2010), there are at least two different ways of attacking an oral argument, and at least two ways of considering and replying to an anticipated counterargument in written discourse. The first strategy consists of a simple (without evidence) challenge to or critique of another person's argument, recognizing at least one objective drawback of that person's approach; the second corresponds to a counter-argument (with evidence) against the conclusion and/or the supporting evidence of another person. Written anticipations may also be of two kinds, namely (a) a simple acknowledgement of a restriction to one's own argument, recognizing at least one objective drawback of one's own approach; and (b) an explicit reply to

an implicit or explicit anticipated counterargument as a strategy for further strengthening one's side. The two sub-types of each category of attack correspond to two different levels of argument quality, the latter (the "b" types) being more dialogic than the former (the "a" types), as they address more effectively the possibility of having alternative arguments and theories for the same issue. We will keep the word "rebuttal" to refer only to the "a" type, renaming the "b" type "underminers."

Integrating the frameworks: Accountability

The assessment of the dimension of accountability presupposes a distinction between the elements of an argument and more importantly the evaluation of the use of evidence. The model that is more used for analyzing argument structure in education is the TAP, which, however, showed crucial problems resulting from the difficulty of distinguishing between data, warrants, and backings. This difficulty was overcome by considering all these different kinds of support to a claim as "grounds" (Clark & Sampson, 2008; Erduran et al., 2004); however, this consideration blurs the fundamental distinction between an argument grounded on evidence and a simple combination of information with conclusions or unbacked arguments.

This problem can be addressed by combining the insights from Kuhn's model with the argumentation schemes theory as advanced by Macagno, Mayweg-Paus, & Kuhn (2015). This proposal advances a unique qualitative criterion for distinguishing information used for supporting a position from the other uses thereof, namely the *functional use of evidence*. For Kuhn and her colleagues, for a piece of information (statistical data, facts, personal examples, etc.) to become evidence, it needs to be functionally used as a support for one's position. On this view, the accountability dimension is assessed considering the relevance one, as better described below. Moreover, according to this approach, a statement becomes a functional support when, and only when, it contains some kind of evidence (Mayweg-Paus, Macagno, & Kuhn, 2016). For Kuhn, this evidence may be either personal (not previously accessible to others) or shared (previously accessible to others). In both cases, what matters is the distinction of evidence from pseudo-evidence, as explained above. In this sense, this mere distinction between functional vs non-functional support is an accountability criterion.

Two practical implications can be drawn from this integration. First, a support is a support only when it contains some type of evidence, and therefore, when it corresponds to Toulmin's backing. Anything else (explanation, elaboration) is part of the argument, not of its support. Second, for a support to be functional, and therefore to be considered as an actual backing, it needs to be *coherent* with the rest of the discourse. To determine coherence, we need to consider Walton's approach to argumentation schemes and dialogue.

Integrating the frameworks:
Relevance and Textuality as aspects of coherence

As we underscored above, a text needs to be a coherent unit in order to be considered as such. Coherence is considered in two complementary ways, as "textuality," or the presence of explicit relations and elements, and as "relevance," or functional connectedness. Clearly, textuality is a manifestation of relevance, or rather a strategy for making the relevance relations clear and available to the interlocutor. For this reason, we distinguish three criteria for assessing the "coherence" of a text.

The first is an indicator of textuality, which we will refer to as "explicitness." If a text makes the relations between its elements explicit, taking for granted (and thus not expressing) only the more obvious and shared relations, it will be perceived as clearer than a text in which the connections are left unstated. When we apply this principle to an argumentative text, we notice that explicitness is a gradual property: we do not need to express everything, and some elements, such as the warrants, are normally and preferentially taken for granted because they are shared (Macagno, 2018). However, when the elements composing an argument are necessary for the identification of the grounds on which a conclusion is based, and the role that it plays for the further conclusion that the text intends to promote, they need to be expressed. Moreover, when a warrant is not necessarily shared, it needs to be expressed, so that it can be clearly identified and, if needed, debated. The lack of explicitness of textual elements and/or the relations between them may result in a straightforward judgment of incoherence due to lack of comprehension, without the analyst even being able to attribute a potential goal, either dialogical or semantic, to the student's contribution. In the case of written discourse, such incoherence is important as it reflects the overall argument quality of a text. The textuality indicator of explicitness can be thus identified with the element that is the guarantee of the cohesion of the text when the relations are not evident, namely the warrant.

The second criterion concerns the relevance of a text, which can be defined as the connectedness of the elements with the goal of discourse (we can call it "Relevance 1"), and the connectedness of the elements among them (we can call it "Relevance 2"). Relevance 1 corresponds to Kuhn's notion of "functionality." However, unlike in Kuhn's account, it is not a specific property of backings, but of all argument elements – conclusions, premises, rebuttals and underminers. Relevance 1, according to Walton (Walton, 2004), can be assessed by determining whether an argument or an argument element is used according to the purpose it is supposed to be used, namely if it can fit an acceptable argument pattern (scheme) leading from it to the conclusion representing its intended and presumable purpose (Macagno, 2019; Macagno & Walton, 2017). This purpose, in the case of persuasive discourse,

corresponds to the dialogic goal of persuading another that one's ideas, positions, evidence are true (acceptable). If an argument/element ignores this goal, it cannot possibly pursue it, and then it cannot be considered as relevant.

Relevance 2 can be evaluated by considering the logical relationship between the argument "elements" (see Table 4). Every argument (i.e. a claim supported by at least one premise) potentially corresponds to a functional argumentation scheme, like the ones presented in Figure 2. Therefore, the logical relevance of the premises used to support a conclusion may be judged considering how well each one "plays its role" as a component of a logical argumentation scheme. If, for example, a proposition is used implicitly as the "warrant" or major premise in an argument from a cause, relevance would be a matter of judging whether the presupposed connection between two events can be considered as a *cause* in a given context. If an event is presented as a consequence of another, its relevance can be assessed by considering whether it is accepted as an effect, and as caused by the first event.

Coding and assessing students' argumentative texts

The frameworks and the integrations thereof for capturing the four DART dimensions can be used for translating the abstract aspects of the quality of an argumentative text into specific proxies that can be identified and measured. Two distinct types of proxies can be distinguished: the indicators of dialogicity, accountability, and textuality, and the criteria for assessing relevance. The first category of proxies is presented in Table 6, which describes six indicators, corresponding to six elements that can be identified in an argumentative text.

Table 6

Argument elements defining quality in students' argumentative texts

Element name	Code	Explanation	Example
Argument	**A**	Any first-level reason (premise and conclusion) given to support one's claim.	I think we should not receive more refugees, **because our country is facing a financial crisis.**
Backing	**B**	Any further back-up to the first-level reasoning, either using personal or shared (socially accessible) information.	I think plastic should be banned because it leads to global pollution (D). **Humans produce tons and tons of plastic per year and only 40% of plastic gets to be recycled.**
Warrant	**W**	Any commonly accepted information that helps clarify/explain the link between data and backing, data and conclusion, and between the conclusion and the goal of the text.	**Animals ingest the plastic and humans eat the animals; therefore**, the humans are eating the plastic. This is why it is said that the plastic is in our food chain (W).
Qualifier	**Q**	Any linguistic modality that reveals some epistemological sophistication or sensibility to generalization, i.e. the fact that one's arguments may not be always valid, but only when certain conditions are met.	I saw a documentary with my mother that said that straws, **along with other plastics**, because they are light, can fly on the wind, they go out to sea, and the penguins **and some other animals** eat them.
Challenge/ limitation to own one's reasoning	**R**	Any limitation or challenge to one's own reasoning integrated as part of it.	(cont. from previous) thinking that they are food and it stays in their bellies, **even if it doesn't kill them (…)**/I think that Portugal should receive more refugees. **Many people think that by allowing refugees into their country, they are also allowing terrorism to enter.**
Underminer	**U**	A reply to an integrated challenge, limitation, or counterargument. Includes meta-dialogical comments or attacks.	There are people who say that we recycle (R). **I see many people not doing it, and even if everybody did, only 40% of the plastic gets to be recycled.**

The dialogicity of an argumentative text can be expressed by the integration of another's viewpoint in one's own argument, or the confrontation of one's viewpoint with another's argument. The code that captures the first aspect is R (rebuttal, defined in a very broad sense following Toulmin's scheme as a challenge/limitation to own one's reasoning), which detects how a text takes into account the strengths of the others' arguments by showing the limits and weaknesses of the advocated position. The second aspect is measured by the code U (underminer), which is an indicator of the writer's analysis and critical assessment of the grounds of the alternative viewpoints, including the relationship between a viewpoint and the accepted definitions, concepts, and common ground.

The accountability dimension of an argumentative text can be measured by considering the persuasive force of its arguments. This variable is captured by the codes A (argument), B (backing), and Q (qualifier). The code A identifies the variety of the reasons used to support a viewpoint: the higher the number of distinct reasons used, the higher the accountability of the contribution. This proxy, however, needs to be considered together with the presence of backing, which measures the evidence used in support of an argument. The code Q expresses the relationship between accountability and dialogicity, as the qualifications to one's conclusion are a sign of the consideration of the defeasible nature of the arguments and the existence of contrary arguments.

The coherence of an argument or another element can be assessed through its relevance and textuality. Textuality can be identified with the presence of the warrants (W) that are necessary for connecting either a premise to the conclusion or an argument to the overall purpose of the text. The presence of the warrants depends on the degree of acceptance and sharedness of the relation – only when the relation cannot be presumed to be obvious to the reader does it need to be made explicit (Anderson, Chinn, Chang, Waggoner, & Yi, 1997; Hitchcock, 1998; Macagno & Capone, 2016; Macagno & Damele, 2013). The absence of warrants leads to the assessment of text relevance. The impossibility or difficulty of finding a relation between an element (A, R, Q) and the goal of the text (Relevance 1) or the other elements (Relevance 2) results in a judgment of irrelevance. In case of backings and underminers, the assessment of relevance refers to the possibility of reconstructing the premises linking them to the arguments they are supporting or attacking. The criteria for determining the relevance of an element are presented in Table 7, which presents definitions and examples of irrelevant elements in students' written discourse.

Table 7
Irrelevant backings and underminers in students' written discourse

Element name	Code	Explanation	Example
Irrelevant Argument	**IA**	The relationship between the premise and the conclusion cannot be presumed to be shared, but it is not expressed as a warrant.	Portugal should not accept refugees because they will bring war and terrorism with them (unaccepted relation taken for granted).
Irrelevant backing	**IB**	Backings whose relationship (W) with the premise to support is not explicit and at the same time not evident. It includes relations that cannot be reconstructed (irrelevant) and also relations that are not semantically valid or generally accepted, even though taken for granted.	1) [Plastic] also pollutes the oceans **as there is a minimum percentage of recycling** (no explicit or self-evident relationship between recycling and polluting the oceans). 2) **Plastic, when it remains in the sun for a long time, "disappears,"** therefore we shouldn't ban it (no clear relationship between "disappearing" and banning). 2) Refugees don't steal jobs from the locals, given that **many of them are children** (no acceptable relationship between children, stealing jobs, and refugees).
Irrelevant underminer	**IU**	Any reply to a counterargument integrated in the line of argument, whose function in the text is problematic because the counterargument it attacks is missing or not made sufficiently explicit.	I think that we should ban the plastic because it is killing many animals. **Humans are irresponsible and do not manage to recycle.** (underminer attacking a counterargument that is implicit and not related to the main argument).

The first two DART dimensions described above (dialogicity and accountability) can be thus translated in a series of proxies that are in turn evaluated through the dimensions of relevance and textuality. A matrix can be thus identified, in which the argumentation aspect is combined with the textual one. The criteria of dialogicity and accountability can be easily

compared with the "logical" criteria developed in argumentation theory for assessing the acceptability of the supports given to a conclusion, namely the number of lines of argument advanced for a conclusion and against the alternative (convergent arguments) and the evidence used for or against an argument (Walton, 2006). Thus, we have two columns representing the lines of argument (collecting the codes referring to the accountability and the dialogicity dimensions related to the other's *viewpoint*) and force (collecting the codes referring to accountability and dialogicity dimensions related to the other's *arguments*).

This matrix is summarized in Table 8. The two columns represent different manifestations of accountability and dialogicity, assessed based on their relevance, considering that irrelevant elements affect not only the comprehension of the single element, but also the understanding or the force of the whole text.

Table 8
An integrated framework for analyzing argumentative texts

DA RT	LINES OF ARGUMENT – Score 1-4	FORCE – Score 1-4
Explanation	Lines of complete arguments (a claim plus a premise) supporting my and/or the other side	Backing vs no backing Anticipation vs no anticipation of counter-arguments
Elements	Strengthen mine or weaken other (A) Qualify (Q) Strengthen other, or limit mine (R)	Backing (B) Underminer (U) Irrelevant backing (IB) Irrelevant underminer (IU)
Quality	Relevant A/R/Q Irrelevant A/R/Q Presence of the necessary W	Relevant B/U Irrelevant B/U Presence of the necessary W

The presence or absence of the indicators of the DA dimensions and the assessment of the quality thereof results in specific scores for the assessment of the quality of argumentative texts (Table 9).

Table 9

An integrated framework for assessing argumentative texts

Scores of DA dimensions	1. Low justification (few As, no Q, no R) 2. Medium justification (several As, no Q, no R / few As, and only one Q or R) 3. Medium justification (several As, at least one R *or* Q) 4. High justification (As, Rs and/or Qs)	**1. Insufficient support:** No arguments are further supported by backings and no attacks are undermined. **2. Weak support:** At least one line of argument is supported by backing(s) (B), *or* at least one possible or actual challenge is identified and undermined (U). **3. Sufficient support:** At least one line of argument is supported by backing(s) (B), *and* at least one possible or actual challenge is identified and undermined (U). **4. Strong support:** All or (in case of many arguments) most of the arguments and counterattacks are supported (B and/or U).
Scores of RT dimensions	Relevant A/R/Q and (needed) W: +1 Relevant B/U and (needed) W: +1 Irrelevant A/R/Q and no (needed) W: -1 Irrelevant B/U and no (needed) W: -1	
Total score: 2–4 Low argument quality level **Total score: 5–6** Medium argument quality level **Total score: 7–8** High argument quality level		

The levels of argument quality are measured based on the following principles:

- Arguments that do not express criticality are considered of low quality (Golder & Coirier, 1994; Kuhn, 2010), even though they are strongly supported (maximum score of uncritical dialogues: 4);
- Argumentative texts of high quality cannot present low levels of relevance and textuality (Macagno, 2016): even if a text is highly accountable and critical (with several backings, arguments, and rebuttals), the presence of irrelevant elements lowers the score to medium;
- A text of medium argumentative quality needs to manifest both dialogicity and accountability with an acceptable degree of relevance. Therefore, if accountability is not expressed through the use of evidence ("force" score: 1), the text needs to show a high level of justification ("lines of argument" score: 4), presupposing the elements constituting dialogicity (total score: 5).

These measures can provide clear guidance on the assessment of argumentative texts, justifying the evaluations thereof based on their twofold textual and argumentative nature. The application of such methods is illustrated through some excerpts from the corpus coded by both authors (Table A1).

This integrated assessment can be used also partially for specific research and pedagogical purposes. The DART dimensions have been designed to combine simplicity (a limited number of codes) with complexity (a high number of dimensions of the same phenomenon). However, some dimensions need a more fine-grained analysis for specific research purposes. For example, accountability can be developed further by considering different levels of backings and underminers, or relevance can be better specified by taking into account the degrees of relevance and the types of relevance (probative relevance vs. pragmatic relevance) (Walton & Macagno, 2016). While the scores and the variables indicated can provide an overall assessment of an argumentative text, the evaluation of a specific dimension thereof can lead to specific sub-codes and a different scoring system.

Conclusion

This theoretical essay intended to propose a method for assessing the quality of argumentative texts considering the four interrelated dimensions of dialogicity, accountability, relevance, and textuality. In our integrated framework, these distinct dimensions are captured by specific proxies. Accountability is measured at two distinct levels: the structural level, consisting of the presence of different lines of argument and rebuttals (originality) and backings (accountability expressed as the use of evidence), and the pragmatic level, considering the logical connection between these elements and the text. Dialogicity is captured by qualifiers, rebuttals, and underminers, elements that address at different levels the interlocutor's viewpoint or potential doubts. Relevance and textuality are conceived as evaluations of the other two dimensions.

The advantages of this proposal consist of the integration of the methods developed in education, text analysis, and argumentation theory to outline a model of assessment that takes into account the complexity of an argumentative text. The limited number of codes considered (6) makes it easy to use and apply, while the analytical depth is guaranteed by the twofold level of analysis (accountability and dialogicity on the one hand, and relevance and textuality on the other).

The limitations of this work are in its theoretical nature. The paper provides a framework for integrating different dimensions of the quality of an argumentative text; however, despite its limited number of codes, the multi-dimensionality can result in interrater reliability issues. Empirical research is needed to test the usability of this method, measuring the agreement between different raters considering different types of argumentative texts. A second empirical issue concerns the validation of the scheme, which needs to be correlated to an independent variable (persuasiveness of the text or achievement of the objectives for which it was written, assessed by a teacher) to evaluate whether it truly and fully captures the argumentative quality of a written text.

Acknowledgements

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Akiguet, S., & Piolat, A. (1996). Insertion of connectives by 9-to 11-year-old children in an argumentative text. *Argumentation*, *10*(2), 253–270.

Alexander, R. (2008). Culture, dialogue and learning: Notes on an emerging pedagogy. In N. Mercer & S. Hodgkinson (Eds.), *Exploring Talk in School: Inspired by the Work of Douglas Barnes* (pp. 91–114). London: Sage Publications.

Anderson, R. C., Chinn, C., Chang, J., Waggoner, M., & Yi, H. (1997). On the logical integrity of children's arguments. *Cognition and Instruction*, *15*(2), 135–167.

Anthony, R., & Kim, M. (2015). Challenges and remedies for identifying and classifying argumentation schemes. *Argumentation*, *29*(1), 81–113.

Aristotle. (1991). Rhetoric. In J. Barnes (Ed.), *The complete works of Aristotle, vol. II.* Princeton: Princeton University Press.

Azar, M. (1999). Argumentative text as rhetorical structure: An application of rhetorical structure theory. *Argumentation*, *13*(1), 97–114.

Cavagnetto, A. (2010). Argument to foster scientific literacy: A review of argument interventions in K–12 science contexts. *Review of Educational Research*, *80*(3), 336–371.

Chinn, C. (2006). Learning to argue. In A. O'Donnell, C. Hmelo-Silver, & G. Erkens (Eds.), *Collaborative learning, reasoning, and technology* (pp. 355–383). Mahwah: Lawrence Erlbaum Publishers.

Choi, Y. H. (1988). Text structure of Korean speakers' argumentative essays in English. *World Englishes*, *7*(2), 129–137.

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.

Coirier, P., & Golder, C. (1993). Writing argumentative text: A developmental study of the acquisition of supporting structures. *European Journal of Psychology of Education*, *8*(2), 169–181.

Cranton, P. (2001). *Becoming an authentic teacher in higher education. Professional practices in adult education and human resource development series*. Malabar: Krieger Publishing Company.

Erduran, S., Ozdem, Y., & Park, J.-Y. (2015). Research trends on argumentation in science education: a journal content analysis from 1998–2014. *International Journal of STEM Education*, *2*(1), 1–12.

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.

Felton, M. (2004). The development of discourse strategies in adolescent argumentation. *Cognitive Development*, *19*(1), 35–52.

Felton, M., & Kuhn, D. (2001). The development of argumentive discourse skill. *Discourse Processes*, *32*(2), 135–153.

Garcia-Mila, M., & Andersen, C. (2007). Cognitive foundations of learning argumentation. In S. Erduran & M. P. Jiménez-Aleixandre (Eds.), *Argumentation in science education* (pp. 29–45). Amsterdam: Springer.

Giddens, A. (1991). *Modernity and self-identity: Self and society in the late modern age*. Stanford: Stanford University Press.

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.

Giora, R. (1998). Discourse coherence is an independent notion: A reply to Deirdre Wilson. *Journal of Pragmatics*, *29*(1), 75–86.

Glassner, A., & Schwarz, B. (2007). What stands and develops between creative and critical thinking? Argumentation? *Thinking Skills and Creativity*, *2*(1), 10–18.

Golder, C., & Coirier, P. (1994). Argumentative text writing: Developmental trends. *Discourse Processes*, *18*(2), 187–210.

Govier, T. (1992). *A practical study of argument*. Belmont: Wadsworth.

Graff, G. (2003). *Clueless in academe: how schooling obscures the life of the mind*. New Haven & London: Yale University Press.

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.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Hanks, W. F. (1989). Text and textuality. *Annual Review of Anthropology*, *18*(1), 95–127.

Hitchcock, D. (1998). Does the traditional treatment of enthymemes rest on a mistake? *Argumentation*, *12*(1), 15–37.

Jimenez-Aleixandre, M. P., Rodriguez, A. B., & Duschl, R. (2000). " Doing the lesson" or "doing science": Argument in high school genetics. *Science Education*, *84*(6), 757–792.

Johnson, R. (2000). *Manifest Rationality: A Pragmatic Theory of Argument*. Hillsdale: Lawrence Erlbaum Associates.

Kelly, G., Druker, S., & Chen, C. (1998). Students' reasoning about electricity: Combining performance assessments with argumentation analysis. *International Journal of Science Education*, *20*(7), 849–871.

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.

Kinneavy, J. (2002). Kairos in classical and modern rhetorical theory. In P. Sipiora & J. Baumlin (Eds.), *Rhetoric and kairos: Essays in history, theory, and praxis* (pp. 58–76). Albany: SUNY Press.

Kuhn, D. (1991). *The skills of argument*. New York: Cambridge University Press.

Kuhn, D. (2010). Teaching and learning science as argument. *Science Education*, *94*(5), 810–824.

Kuhn, D., & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents' thinking. *Psychological Science*, *22*(4), 545–552.

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., & Park, S.-H. (2005). Epistemological understanding and the development of intellectual values. *International Journal of Educational Research*, *43*(3), 111–124.

Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development*, *74*(5), 1245–1260.

Leech, G. (1983). *Principles of pragmatics*. London: Longman.

Leitão, S. (2000). The potential of argument in knowledge building. *Human Development*, *43*(6), 332–360.

Levinson, S. (2012). Action formation and ascription. In J. Sidnell & T. Stivers (Eds.), *The handbook of conversation analysis* (pp. 101–130). Chichester: John Wiley & Sons, Ltd.

Long, M. (1996). Authenticity and learning potential in L2 classroom discourse. *University of Hawai'i Working Papers in English as a Second Language 114*(2), 127–149.

Macagno, F. (2016). Argument relevance and structure. Assessing and developing students' uses of evidence. *International Journal of Educational Research*, *79(1)*, 180–194.

Macagno, F. (2018). Assessing relevance. *Lingua*, *210–211*, 42–64.

Macagno, F. (2019). Coding relevance. *Learning, Culture and Social Interaction*, Advance online publication.

Macagno, F., & Capone, A. (2016). Uncommon ground. *Intercultural Pragmatics*, *13*(2), 151–180.

Macagno, F., & Damele, G. (2013). The dialogical force of implicit premises. Presumptions in enthymemes. *Informal Logic*, *33*(3), 365–393.

Macagno, F., & Konstantinidou, A. (2013). What students' arguments can tell us: Using argumentation schemes in science education. *Argumentation*, *27*(3), 225–243.

Macagno, F., Mayweg-Paus, E., & Kuhn, D. (2015). Argumentation theory in education studies: Coding and improving students' argumentative strategies. *Topoi*, *34*(2), 523–537.

Macagno, F., & Walton, D. (2015). Classifying the patterns of natural arguments. *Philosophy and Rhetoric*, *48*(1), 26–53.

Macagno, F., & Walton, D. (2017). *Interpreting straw man argumentation. The pragmatics of quotation and reporting.* Amsterdam: Springer.

Manz, E. (2016). Examining evidence construction as the transformation of the material world into community knowledge. *Journal of Research in Science Teaching*, *53*(7), 1113–1140.

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. (2016). Developing argumentation strategies in electronic dialogs: Is modeling effective? *Discourse Processes*, *53*(4), 280–297.

Metaxas, N., Potari, D., & Zachariades, T. (2016). Analysis of a teacher's pedagogical arguments using Toulmin's model and argumentation schemes. *Educational Studies in Mathematics*, *93*(3), 383–397.

Michaels, S., Connor, C., & Resnick, L. (2008). Deliberative discourse idealized and realized: accountable talk in the classroom and in civic life, *27*(4), 283–297.

Newell, G. E., Beach, R., Smith, J., & Van Der Heide, J. (2011). Teaching and learning argumentative reading and writing: A review of research. *Reading Research Quarterly*, *46*(3), 273–304.

Nussbaum, M. (2003). Appropriate appropriation: Functionality of student arguments and support requests during small-group classroom discussions. *Journal of Literacy Research*, *34*(4), 501–544.

Nussbaum, M. (2008). Using argumentation vee diagrams (AVDs) for promoting argument-counterargument integration in reflective writing. *Journal of Educational Psychology*, *100*(3), 549–565.

Nussbaum, M. (2011). Argumentation, dialogue theory, and probability modeling: Alternative frameworks for argumentation research in education. *Educational Psychologist*, *46*(2), 84–106.

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.

Nussbaum, M., & Schraw, G. (2007). Promoting argument- counterargument integration in students' writing. *The Journal of Experimental Education*, *76*(1), 59–92.

O'Keefe, D. (1977). Two concepts of argument. *Journal of the American Forensic Society*, *13(*3), 121–128.

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, *328*(5977), 463–466.

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.

Paglieri, F. (2015). Bogency and goodacies: On argument quality in virtue argumentation theory. *Informal Logic*, *35*(1), 65–87.

Perkins, D., Farady, M., & Bushey, B. (1991). Everyday reasoning and the roots of intelligence. In J. Voss, D. Perkins, & J. Segal (Eds.), *Informal reasoning and education* (pp. 83–105). Hillsdale: Lawrence Erlbaum Associates.

Rapanta, C., Garcia-Mila, M., & Gilabert, S. (2013). What is meant by argumentative competence? An integrative review of methods of analysis and assessment in education. *Review of Educational Research*, *83*(4), 483–520.

Rapanta, C., & Macagno, F. (2016). Argumentation methods in educational contexts: Introduction to the special issue. *International Journal of Educational Research*, *79(1)*, 142–149.

Rapanta, C., & Walton, D. (2016a). Identifying paralogisms in two ethnically different contexts at university level/Identificación de paralogismos en dos contextos universitarios diferenciados étnicamente. *Infancia y Aprendizaje*, *39*(1), 119–149.

Rapanta, C., & Walton, D. (2016b). The use of argument maps as an assessment tool in higher education. *International Journal of Educational Research*, *79(1)*, 211–221.

Reinhart, T. (1980). Conditions for text coherence. *Poetics Today*, *1*(4), 161–180.

Reznitskaya, A., & Wilkinson, I. (2015). Professional development in dialogic teaching: Helping teachers promote argument literacy in their classrooms. In D. Scott & E. Hargreaves (Eds.), *The SAGE handbook of learning* (pp. 219–232). London: SAGE Publications Ltd.

Sampson, V., & Clark, D. (2006). Assessment of argument in science education: A critical review of the literature. In S. Barab, K. Hay, & D. Hickey (Eds.), *Proceedings of the 7th international conference on learning sciences* (pp. 655–661). Bloomington: ACM Press.

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.

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.

Schank, R., Collins, G., Davis, E., Johnson, P., Lytinen, S., & Reiser, B. (1982). What's the point? *Cognitive Science*, *6*(3), 255–275.

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.

Stab, C., & Gurevych, I. (2014). Identifying argumentative discourse structures in persuasive essays. In A. Moschitti, B. Pang, & W. Daelemans (Eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 46–56). Doha: Association for Computational Linguistics.

Toulmin, S. (1958). *The uses of argument.* Cambridge: Cambridge University Press.

Toulmin, S., Rieke, R., & Janik, A. (1984). *An introduction to reasoning.* New York: Macmillan Publishing Company.

Van Dijk, T. (1977a). Sentence topic and discourse topic. *Papers in Slavic Philology*, *1*, 49–61.

Van Dijk, T. (1977b). Pragmatic macrostructures in discourse and cognition. In M. De Mey, R. Pinxten, M. Poriau, & F. Vandamme (Eds.), *CC77. International Workshop on the Cognitive Viewpoint* (pp. 99–113). Ghent: University of Ghent Press.

Van Dijk, T. (1980). *Text and context explorations in the semantics and pragmatics of discourse.* London & New York: Longman.

Van Dijk, T., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

van Eemeren, F., & Grootendorst, R. (1984). *Speech acts in argumentative discussions: A theoretical model for the analysis of discussions directed towards solving conflicts of opinion.* Dordrecht: Floris Publications.

van Eemeren, F., & Grootendorst, R. (2004). *A systematic theory of argumentation: The pragma-dialectical approach.* Cambridge: Cambridge University Press.

Van Lier, L. (1996). *Interaction in the language classroom: Awareness, autonomy and authenticity.* London: Longman.

Vatz, R. (1973). The myth of the rhetorical situation. *Philosophy & Rhetoric*, *6*(3), 154–161.

Von Aufschnaiter, C., Erduran, S., Osborne, J., & Simon, S. (2008). Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge. *Journal of Research in Science Teaching*, *45*(1), 101–131.

Walton, D. (1989). *Informal logic*. New York: Cambridge University Press.

Walton, D. (1990). What is reasoning? What is an argument? *Journal of Philosophy*, *87(*8), 399–419.

Walton, D. (1992). *Plausible argument in everyday conversation*. Albany: State University of New York Press.

Walton, D. (1995). *Argumentation Schemes for Presumptive Reasoning*. Mahwah: Routledge.

Walton, D. (2004). *Relevance in argumentation*. Mahwah: Erlbaum.

Walton, D. (2006). *Fundamentals of critical argumentation*. New York: Cambridge University Press.

Walton, D. (2010a). Objections, rebuttals and refutations. In J. Ritola (Ed.), *Proceedings from the 8th Conference of the Ontario Society for the Study of Argumentation, June 3–6, 2009, University of Windsor* (pp. 1–11). Windsor: OSSA.

Walton, D. (2010b). Why fallacies appear to be better arguments than they are. *Informal Logic*, *30*(2), 159–184.

Walton, D., & Macagno, F. (2016). Profiles of dialogue for relevance. *Informal Logic*, *36*(4), 523–556.

Walton, D., Reed, C., & Macagno, F. (2008). *Argumentation schemes*. New York: Cambridge University Press.

Witte, S. P., & Faigley, L. (1981). Coherence, cohesion, and writing quality. *College Composition and Communication*, *32*(2), 189–204.

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.

Corresponding authors

Fabrizio Macagno
ArgLab – Reasoning and Argumentation Lab, Faculty of Social Sciences and Humanities, Universidade Nova de Lisboa, Lisbon, Portugal
E-mail: fabrizio.macagno@fcsh.unl.pt

Chrysi Rapanta
ArgLab – Reasoning and Argumentation Lab, Faculty of Social Sciences and Humanities, Universidade Nova de Lisboa, Lisbon, Portugal
E-mail: chrysi.rapanta@fcsh.unl.pt