System 101 (2021) 102576

Contents lists available at ScienceDirect

System

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

The effects of in-class debates on argumentation skills in second language education

Abid el Majidi [a,*], Daniel Janssen [b], Rick de Graaff [c]

[a] Utrecht University, Trans 10, 3512 JK, Utrecht, the Netherlands
[b] Utrecht University, Trans 10/office 0.43a, 3512 JK, Utrecht, the Netherlands
[c] Utrecht University, Trans 10/office 1.60, 3512 JK, Utrecht, the Netherlands

ARTICLE INFO

Keywords:
In-class debates
Second language
Argumentation skills
Secondary school
Written and oral argumentation
Argumentation pedagogy

ABSTRACT

The skill of argumentation in the second language (L2) is problematic for secondary school students who are deficient in expressing structurally and qualitatively appropriate arguments. Debate has been widely acclaimed as an effective pedagogical tool that can improve L2 argumentation skills of students. However, the existing evidence is anecdotal. This intervention study, which employed a pretest-posttest control group design involving eight classes at three secondary schools in The Netherlands, was conducted to investigate to what extent debate instruction in L2 affects a number of L2 structural (e.g. sub-arguments and rebuttals) and quality (e.g. elaboration and persuasiveness of arguments) aspects of argumentation. To gauge the effect of the intervention, we analyzed written and oral opinion tasks produced by the participants at the beginning and towards the end of the intervention. We used an adjusted version of Toulmin's argumentation model to undertake the structural analysis and a 5-point scale rubric to assess different aspects of reasoning quality. Multilevel analysis of the data revealed that debate instruction had a positive effect upon a number of structural components and quality aspects of the written and oral argumentation skills. These findings led us to conclude that debate constitutes a conducive pedagogy for honing argumentation skills.

1. Introduction

The ability to reason critically is arguably one of the most fundamental skills underlying success in everyday life as well as academia and professional careers (Butt, 2010; Yeh, 1998). Performing these roles successfully "requires the ability to understand, decide, or persuade effectively, either verbally or in writing, through the process of argumentation" (Iordanou, 2013, p. 292). The ability to generate and evaluate arguments has been also widely recognized as a key indicator of good critical thinking ability (Mercier, 2011; Mercier & Sperber, 2011). Good thinking skills enable people to critically verify incoming information, consider alternative viewpoints and produce counterarguments (Osana & Seymour, 2004). Roy and Macchiette (2005) point out that critical thinking stimulates adopting a positive attitude and enables going beyond challenging an opposing view in an attempt to develop new understanding, think up new theory and make speculations about the future.

Many studies have voiced concerns about students' argumentative abilities in the first Language (L1) (e.g. Chen et al., 2016; Crowell & Kuhn, 2014) as well as in the second/foreign language (L2/FL) (e.g. Huh & Lee, 2014; Qin & Karabacak, 2010). Empirical

* Corresponding author.
E-mail addresses: a.elmajidi@uu.nl (A. Majidi), D.M.L.Janssen@uu.nl (D. Janssen), r.degraaff@uu.nl (R. de Graaff).

https://doi.org/10.1016/j.system.2021.102576
Received 5 October 2020; Received in revised form 9 June 2021; Accepted 30 June 2021
Available online 1 July 2021

evidence shows that students struggle with providing adequate justifications to their claims (Hsu et al., 2015) and generating counterarguments to rebut the opposing side (Liu & Stapleton, 2020; Sadler, 2004; Stapleton & Wu, 2015). Weaknesses in argumentative reasoning can have serious implications for success in academia and professional careers and therefore warrants nurturing (Crowell & Kuhn, 2014; Kuhn, 1991; Kuhn & Udell, 2003; Yeh, 1998). Or in the words of Yeh (1998): the capacity of producing "effective arguments influences grades, academic success, and preparation for college and employment" (p. 49).

A number of studies have linked deficiencies in argumentation skills (both in L1 and L2 context) to the lack of adequate and effective instruction rather than to the inherent inability of students to acquire such skills (Butt, 2010; Hirvela, 2017; Reznitskaya el al., 2001; Walker & Kettler, 2020). Osana and Seymour (2004) maintain that fostering and evaluating students' ability to make critical judgements and evaluations is a crucial responsibility that needs to be assumed by the educational system. Therefore, seeking pedagogical tools that promote students' ability to reason with critical thinking and argue with credible evidence is an urgent overarching priority. The good news is that students, including "EFL novice writers can, in an appropriate learning context, demonstrate progress in formulating effective arguments" (Cheng, 2010, p.140).

There are two notable variables that can interfere with argumentation performance in L2: culture and language proficiency. Culture is a variable that may play a role in the development of argumentation skills. In some non-western societies like countries dominated by Confucianism culture, teachers are viewed as authority figures who cannot be challenged or criticized by students (Heimgärtner, 2013). However, in most western countries (the context of this study), which are characterized by small power distance societies as teachers and students relatively share more equal power distribution, students are stimulated to speak up and can even challenge teachers (Hsu et al., 2017). In such a context, the culture variable is less likely to interfere with argumentation expression and development.

Limited L2 proficiency, on the other hand, may also interfere with students' ability to express their arguments effectively in L2. Some studies (e.g. Cheng & Chen, 2009; Hsu et al., 2017) suggest that limited linguistic resources may hinder students from properly developing and justifying their arguments and employing certain sophisticated components of arguments. Though this may apply to a number of aspects of argumentation, as demonstrated in the study of Cheng and Chen (2009) some aspects of argumentation do not seem to be affected by language proficiency, such as "handling oppositional structures" (p. 23). Cheng and Chen (2009) call for "additional research … to determine the role that L2 proficiency, cognition and culture play in students' use of argument structures" (p. 45).

In the Dutch context, students' argumentative skills also receive inadequate attention, especially in secondary education, despite their noted importance (Van Eemeren et al., 2015). L2 argumentative skills are of paramount importance for both secondary school and university classrooms in the Netherlands. In secondary schools, students are required to compose argumentative L2 essays, employing well-reasoned arguments. In addition, an increasing number of Dutch students take internationally recognized tests, such as Cambridge ESOL in which oral and written argumentation abilities are assessed. Furthermore, in Dutch bilingual schools many subjects (e.g. history) that strongly involve argumentation are instructed in English. As to the university context, the majority of Dutch students need to write many L2 essays and papers which involve argumentation in one way or another (De Haan & Van Esch, 2005).

In-class debate is regarded as a potentially effective pedagogical tool that may help to improve learners' reasoning/argumentative skills (Malloy et al., 2020; Oros, 2007; Zorwick & Wade, 2016). The debate process is believed to offer one of the best mechanisms for operationalizing and applying the principles of critical thinking (Butt, 2010; Freeley & Steinberg, 2005; Roy & Macchiette, 2005). Its environment contains many reasoning incentives that hone argumentation skills. Involvement in debate forces students to search, inspect and evaluate arguments, overcome personal prejudices and biases, identify inconsistencies and inadequacies in opponents' line of reasoning and eventually engineer well-thought and persuasive arguments.

Though there is a wide recognition that debate may promote argumentation skills, there is no empirical evidence available yet to support this assumption in the L2 context. To fill this gap, we conducted an intervention study in which we investigated the effects of in-class debates on the development of argumentation skills of Dutch secondary school students in English classes by identifying and comparing developmental features (structural and quality aspects) of argumentative skills in samples of their written and oral production produced by an intervention and a control group.

2. Theoretical grounds

The potential of in-class debates for argumentation skills development can be motivated from multiple theoretical perspectives. One theoretical perspective comes from dialogic argumentation, with its roots in the everyday social practice of talk. Dialogic argumentation is regarded as an essential pathway for fostering individual argumentative reasoning (Crowell & Kuhn, 2014; Kuhn, 2018; Michaels et al., 2008). Along the same lines, Reznitskaya et al. (2001) argue that reasoning is fundamentally dialogical and, therefore, its development is best fostered in social dialogical settings. Dialogic argumentation involves students in social negotiations that enable them to gain insights into the strengths/weaknesses of their arguments and hence improve them (Chen et al., 2016). Through social interaction students become not only exposed to alternative perspectives, but they get engaged in argumentative interaction which enables them to compare their arguments with each other, notice gaps in them and hence attend to and remedy the problematic areas in their reasoning. Under such circumstances and following Long's (1996) interaction theory –which theorizes that through interaction L2 learners become aware of the shortcomings in their input and thus modify their output –we posit that students go through argumentation processing that stimulate drawing attention to reasoning gaps and hence reflect on and revise their argumentation skills. This assumption is supported by Newell et al. (2011) who contend that "students acquire argumentative literacy practices through active participation in dialogic interactions" (p.292). The competitive atmosphere of debate is likely to fuel these interactions between students and accordingly sharpen up their reasoning abilities (e.g. Roy & Macchiette, 2005).

Another theoretical perspective has to do with the conceptualization of argumentation: *learning to argue* and *arguing to learn* (Hirvela, 2017). In the *learning to argue* conceptualization, students are taught how to argue through teaching them the components of argumentation; in other words, arguing is seen as an end in itself. On the other hand, *arguing to learn* perspective conceptualizes argumentation as a "means by which we rationally resolve questions, issues, and disputes and solve problems" (Jonassen & Kim, 2010, p. 439). According to Hirvela (2017), the *arguing to learn* orientation empowers learners to think beyond the argument structure and fosters their analytical and critical thinking skills. Debating may subscribe to *the arguing to learn* orientation. In debates, argument is perceived as a medium through which a functional goal can be attained, and that is defending one's position and weakening that of their opponents. This orientation stimulates students to see argument as "process, not product" (Hirvela, 2017, p. 72).

More evidence to support our premise that debate may be propitious for argumentation skills comes from the potential effectiveness of involving students in argumentative activities that connect speaking/talking and writing. Chen et al. (2016) maintain that engaging students in dialogical interactions and writing constitute a promising avenue for improving their argumentative competence. The argumentative literacy practices students gain from active participation in dialogic interactions are likely to transfer to argumentative writing (Newell et al., 2011). Reznitskaya et al.'s (2001) empirical work substantiated this assumption. In this study, the students who participated in discussions involving controversial issues wrote essays that showed a greater number of arguments, counterarguments, rebuttals, uses of formal argument devices and references to text information than the essays of the control students who did not participate in discussions. Reznitskaya et al. concludes that "reasoning skills acquired in discussion transferred to a different context, from collaborative oral discussions to the individual task of persuasive writing" (p. 171).

The audience-centered approach to argumentation that characterises debate presents another relevant perspective that supports our hypothesis. Many studies showed that audience awareness pushes debaters to hone their reasoning skills. For example, Chen et al. (2016) point out that when "students [are] expected to be challenged and critiqued by their peers [they push] pushed themselves to refine and reshape their argument to be convincing" (p.130). In the same vein, Midgette, Haria and MacArthur's (2008) study revealed that students with audience-oriented goals were more likely to consider opposing positions and rebut them. Even during preparation (e.g. case[1] writing), audience awareness can affect the construction of arguments, in that arguers consciously think about the audience's objection and response and accordingly reflect upon and refine their line of reasoning (Johnson, 2013). The last theoretical perspective has to do with students' attitude towards debate. Students' favorable attitude towards debating is likely to lead to high cognitive engagement in the learning process, thereby leveraging the learning potential offered by the debate experience (el Majidi et al., 2015, 2018).

3. Debate-argumentation research

Several studies reported improvement in the argumentative skills of students that took part in debates. The majority of these studies were conducted in L1 and were based on students' self-reports and instructors' observations. For example, Zorwick and Wade (2016), whose study was based on analysis of student course evaluations and on observations of the authors and other instructors, reported that there was unanimous agreement among Social Studies/History teachers (participants) that debate activities enhanced their students' ability to "write arguments to support claims using valid reasoning and relevant and sufficient evidence" (p. 441). In another study involving American university students, Oros (2007) reported that the students who participated in the *Introduction to World Politics* course (with debates) produced far more arguments and with a higher degree of support in the final essays than the students who participated in the *Introduction to Political Science* course (without debates). Oros attributed this difference in part to the debate experience. He also notes that this difference persisted after the conclusion of the courses.

Studies that have investigated the argumentation-debate relationship in the L2/FL context are very scarce. These studies predominantly drew on anecdotal evidence, and they mainly elicited data through questionnaires and interviews (e.g. Gulnaz, 2020; Zare & Othman, 2015). These studies, which have massively recognized that debate promotes critical reasoning, lack empirical evidence that unequivocally establishes a causal link between debate and argumentation skills. Therefore, the current study is needed to fill this lacuna by providing empirical evidence about the extent to which debate pedagogy can affect L2 argumentation skills.

It is noteworthy that, in general, the studies that investigated L2/FL argumentative performance are scant (Paek & Kang, 2017; Qin & Karabacak, 2010). These studies revealed that L2 learners' arguments were feeble and structurally simple. Their papers, for example, mainly contained two basic elements of argument structures, namely, claims and data; they hardly included sophisticated structural elements, such as counterarguments and rebuttals, "which may make the arguments less persuasive and lower the quality of writing" (Paek & Kang, 2017, p. 117).

Hirvela (2017) regards the negligence of argument in L2 research agenda (especially in L2 writing) odd, unjustifiable and unacceptable seeing its manifest importance. For example, L2 argumentative competence is viewed as an important indicator of L2 writing ability "as argumentation is at the heart of SL [second language] writing assessment" (Hirvela, 2017, p. 69). In addition, L2 argumentation may also boost L2 language development. Chapple and Curtis (2000) hold that improvement in L2/FL analytic and critical thinking may lead to improvement in L2/FL language skills. Similarly, Pally (1997) contends that "complex, synthetic reasoning needs to be practiced in the L2 in order for students to master -and challenge- L2 language and argument" (p. 299).

[1] A case is "a cohesive set of [written] arguments [prepared beforehand] that justify the side of the topic that they have been assigned" (Snider & Schnurer, 2006, p. 26).

4. This study

To recapitulate, in spite of the importance of L2 argumentation skills, "few studies have addressed the pedagogical needs of developing explicit instructional approaches to foster argumentation skills for L2 writers" (Cheng, 2010, p. 120). The current study endeavours to fill part of this research gap by investigating the effects of a debate intervention in L2 on students' L2 oral and written argumentation skills by analyzing samples of their written and oral output. Learners' "verbal and written arguments are likely to differ in complexity" (Berland & McNeill, 2010, p. 790). Therefore, investigating both oral and written arguments is likely to provide a complete picture of the impact of debate on the argumentative competence of L2 learners. Many researchers (e.g. Kathpalia & See, 2016; Sampson & Clark, 2008; Stapleton & Wu, 2015) stress the need to examine both the structural components and quality aspects of the produced arguments to obtain a reliable picture of the quality of reasoning in learners' argumentative competence. As Stapleton and Wu (2015) argue "the surface structure, or shell of the argument, may appear appropriate or even exemplary, but the actual substance could still be exceedingly weak'' (p. 12). It is thus imperative to utilize an integrated argumentation assessment framework that takes both the structural (e.g. data, warrants, rebuttals, etc.) and the quality (the substance) (e.g. clarity, elaboration and persuasiveness of arguments) aspects of the argument into account. This study addressed the following research questions:

1. What is the effect of in-class debates on the structural components of the written and oral arguments produced by Dutch secondary school students?
2. What is the effect of in-class debates on the quality of the written and oral arguments produced by Dutch secondary school students?

Hypotheses

The theoretical considerations (discussed above) and previous research (e.g. Oros 2009; Zorwick & Wade, 2016) led us to anticipate:

1. That students engaged in the debate intervention would produce more written and oral arguments and that these arguments would include more sophisticated structural patterns.
2. That students engaged in the debate intervention would produce more written and oral arguments that exhibit better reasoning quality.

5. Method

5.1. Design

To answer these research questions, an intervention with a pretest-posttest control group quasi-experimental design was conducted. The intervention group participated in ten debates (one debate per week) which were part of the class curriculum, with each lasting approximately 50 min. It is noteworthy that the intervention group did not receive instruction about the structural and quality aspects of argumentation. To enhance the external validity of the findings, we gathered data from three different secondary schools. The participating teachers were instructed and familiarized with the content of the intervention by the first author. The data for this study consisted of written and oral opinion tasks (in which the intervention and control students argued for/against a side of a controversial topic, see 5.4 for more elaboration) and were elicited at the beginning (pretest) and towards the end (posttest) of the intervention.

5.2. Participants

The study's sample consisted of eight intact classes at three secondary schools in The Netherlands ($n = 147$). Five classes were in their fifth year of higher general secondary education ($n = 89$) and three classes were in their fourth year of pre university secondary education[2] ($n = 58$). Five classes served as the intervention group ($n = 96$) and three as the control group ($n = 51$). The participants included 88 females and 59 males, ranging in age from 15 to 18. The English proficiency level of all classes spanned mostly the B1[3] and B2[4] levels as estimated by their teachers. With the exception of one intervention group that received on average two English sessions of 50 minutes per week, other groups received three sessions of 50 minutes. Both groups received regular instruction consisting of activities dealing with the four language skills. For the purpose of this study, while the intervention students were involved in the debate intervention (once a week), the control students received extra regular instruction (in which the four skills were further practiced). See the intervention section for details about argumentation.

[2] For more information about the Dutch education system see: https://www.nuffic.nl/en/subjects/education-in-the-netherlands/#secondary-education.

[3] Intermediate level according to the Common European Framework of Reference (CEFR).

[4] Upper-intermediate level according to CEFR.

5.3. The intervention

The debate task was developed and validated in a previous study following the principles and guidelines of educational design research (see, for example, McKenney & Reeves, 2012). The intervention students who participated in ten debates were informed at least one week in advance of the debate topic. They were allowed to choose the topics that interested them and the side they wanted to defend (i.e. an affirmative or a negative side). We employed two debate formats that were extensively tested before the intervention took place: debating in a group of four debaters (two students in favor and two against) and a one-to-one debating format. All debates had three phases: constructive speech, rebuttal and clash (see e.g. Snider & Schnurer, 2006).

Table 1 presents the activities performed in each debate session in addition to the tasks involving argumentation performed by the control students during the intervention.

Since students in the intervention group (in different schools) were allowed to choose topics, their readings were not exactly the same. This also applies to the students in the control group who were allowed to choose their own articles (which as indicated in Table 1 included argumentative articles). It follows, then, that specific content is not likely to interfere with our findings. It is worth mentioning that the intervention students also wrote persuasive essays and letters. However, in order not to disadvantage the control students in terms of the amount of practice with the argumentative discourse, we allowed them to write more essays and letters. More precisely, during the intervention the control group wrote on average two more essays and one more letter than the intervention group. Additionally, the control groups read and summarized three more articles, of which some were argumentative.

5.4. Procedures

To examine whether the debate intervention had impact on written and oral arguments, we analyzed and compared two written and two oral texts (argumentative opinion tasks) produced by the intervention and control groups during pre- and posttest (see Table 2). As to the writing task, we selected two controversial topics: (1) capital punishment should be legalized, and (2) abortion should be banned. Topics like capital punishment and abortion compared to many other social issues are accessible topics with clear sides (with each side having abundant arguments). Having clear sides is likely to enable our participants to generate counterarguments and rebuttals. These topics were previously piloted and proved to fit our context.

To evaluate the students' oral arguments, we employed (semi) unplanned opinion tasks, involving different topics (e.g. smoking should be banned) which were randomly assigned to students. The students received 7 minutes to prepare and write down any notes they wished to use in their oral opinions. Not offering students 7 minutes of preparation would lead to very short speeches which would not reflect the oral argumentative competence of the participants. The conditions around pre- and post-assessments were the same for both the intervention and control groups. School and parental permission forms were obtained prior to the beginning of the intervention.

It is important to note that during the 25 minutes of preparation time, the participants received a preselected article from the instructor with opposing views on the same topic and were allowed to surf the net for more arguments to build their content knowledge. 25 minutes of preparation time was deemed to be sufficient for the participants to prepare for the writing opinion task (Qin & Karabacak, 2010). After preliminary preparation, we collected the texts to prevent the students from copying the articles. We then tasked the students to write a 15-minutes timed text defending their side.

We believe that a free writing opinion task which is not governed by any particular writing conventions would allow students to focus more on arguments and their articulation.[5] Furthermore, we believe that free opinion tasks are likely to provide a more fine-grained picture of students' use of counterarguments and rebuttals than essays since our students were already instructed (before the intervention) to compose two-sided essays. The topics of the pre- and post-test opinion tasks (capital punishment and abortion) were counterbalanced to avoid any potential topic effect.

5.5. Data analysis

5.5.1. Argument structure analysis

A model that has been widely accepted and used as an instrument of structural analysis of arguments is Stephen Toulmin's model of argumentation (Huh & Lee, 2014; Xargia, 2016). Toulmin's analytical framework (1958, 2003) has been extensively used in L1 context and is increasingly adopted to examine learners' argumentation in L2/FL context (Cheng & Chen, 2009; Huh & Lee, 2014).

Toulmin's argument model involves a structural analysis that breaks an argument into six components which are then divided into two groups. The first group, known as primary elements, form the foundation for the argument: claim (i.e. the conclusion of an argument or position being argued for), data (i.e. the evidence advanced to support a claim) and warrant (i.e. the reasoning that establishes a link between data and claim). Warrants are mostly implicit as the link is expected to be understood. The second group of elements, which are optional, are known as secondary elements and can be used to strengthen the argument by complementing the primary elements. These elements are rebuttal (i.e. addresses the conditions which could defeat the claim), qualifier (i.e. placing limits on the strength of the claim) and backing (i.e. support for the warrant).

[5] This is based on our experience and piloting stage. Unlike, for example, essays which are governed by a number of conventions (e.g. structure, thesis statement, etc.), free opinion tasks, which are not constrained by these conventions, allow students to direct their attention more to the construction and formulation of arguments.

Table 1
Main activities conducted during the intervention.

Intervention students	Control students
Pre-debate stage • Reading two articles (related to the topic under debate) and summarizing them. The instructor provided one article and the students had to find another one. • Writing cases During-debate stage • Noting down the arguments of the opponents as debaters have to rebut them during the rebuttal and clash stages.	• Reading and summarizing (argumentative) newspaper articles (e.g. the Guardian) covering current issues, including political and changes in policies issues. • Writing controversial persuasive essays (e.g. school uniforms should be introduced in schools). • Writing letters, especially complaint letters in which students have to express dissatisfaction with a particular service and accordingly convince the addressed company of providing a refund.

Table 2
Data collection procedures.

Pretest	
• Writing a 15-minute opinion task after 25 minutes preparation time. • Producing an oral opinion (with no time limits) after 7 minutes of preparation time.	
Intervention (intervention group)	No intervention (control group)
See Table 1 for the activities conducted during the intervention	See Table 1 for the activities conducted during the intervention
Posttest	
• Writing a 15-minute opinion task after 25 minutes preparation time. • Producing an oral opinion (with no time limits) after 7 minutes of preparation time.	

Although there are some criticisms levelled against Toulmin's model, most of which mainly concern the difficulties of discerning each element, several studies have demonstrated a close relationship between the Toulmin elements and argumentative writing qualities (e.g. Crammond, 1998; Paek & Kang, 2017).

In order to deal with the inadequacies of the Toulmin model, several researchers modified this model, mostly by reducing or modifying its features (Cheng & Chen, 2009; Connor, 1990; Crammond, 1998). Accordingly, to serve the purpose of this research, we also modified Toulmin's model to accommodate the range of structural variations that are encountered in the arguments composed by L2 secondary school students in our study.

In the current study (see appendix A), we coded the Toulmin's primary elements (i.e. claim, data and warrant) and following Crammond (1998) and Cheng and Chen (2009), we added four structural elaborations and modifications to Toulmin's model: Firstly, the backing element in the present study includes in addition to backing for warrants also backing for data, qualifiers, alternative solutions, counterarguments and rebuttals. Secondly, qualifiers are modified to include the conditions that limit the applicability or validity of the claim (reservation) or the conditions under which the claim would apply (constraints). Thirdly, unlike in the Toulmin model, rebuttals in the present study refer to the debater's response to opposing views that challenge the debater's claim. Fourthly, following Crammond, and Cheng and Chen, we also coded alternative solutions which refer to possible solutions or answers to the issue under consideration. Lastly, after we had performed some preliminary coding, we realized that it was needed to include an additional element also employed by Karabacak and Qin (2016): background information which consists of general information about the issue and to introduce a novel element in our analysis: subordinate arguments which are sub-arguments that back up a main argument.

Lastly, to assess the overall complexity and structural sophistication of the argumentation presented in each text, we developed a classification scheme (see appendix B) that classifies texts in terms of the structural complexity of argumentation pattern. In developing this framework, we drew on Osborne et al. (2004), Venville and Dawson (2010) and Erduran et al. (2004). Of particular importance in this framework is that texts consisting of rebuttals are accorded a high score because, as Kuhn (1991) argued, the ability to use rebuttals is ''the most complex skill,'' as an individual must ''integrate an original and alternative theory, arguing that the original theory is more correct'' (p. 145). Thus, "rebuttals are an essential element of arguments of better quality and demonstrate a higher-level capability with argumentation" (Osborne et al., 2004, p. 1009).

To perform structural analysis, all texts received an overall score for argumentation complexity and a frequency score for each dependent variable (e.g. data and sub-arguments). For example, the sub-argument score reflects the number of sub-arguments used in each text (see Appendix D).

5.5.2. Argument quality analysis

The adapted Toulmin argumentative framework used for evaluation in the present study only addresses the surface argumentative structure, the shell of the argument (see Qin & Karabacak, 2010). However, for a comprehensive evaluation of the argumentative competence, the quality of the student-generated arguments needs to be assessed as well.

To assess the quality of each production, we developed a 5-point rating scale comprising a set of items that tap into different aspects of the reasoning quality, including organization (e.g. to what extent are the arguments well organized and flow well), sufficiency, clarity, elaboration, relevance, persuasiveness and addressing the opposing view in addition to holistically assessing the overall quality of each production (see Appendix C for more elaboration). Following this model, arguments increase in quality as they become

increasingly organized, sufficient, clear, elaborated, relevant, persuasive and address the opposing view. In developing this quality-oriented evaluation framework, we drew heavily on coding schemes developed and validated by Qin and Karabacak (2010) and Ferretti et al. (2000). In appendix D, we have provided an Example of the scoring procedures for two written texts.

5.6. Interrater reliability

To perform a structural analysis of the written and oral arguments produced, we identified elements of arguments and calculated their frequencies. Efforts to achieve acceptable interrater reliability were made as follows: The first two authors first went through practice sessions, scoring some texts and then compared scores and discussed the discrepancies to clarify and standardize/calibrate the interpretation of coding. After the practice sessions, the first author coded the entire dataset. To calculate reliability scores, a random sample of 25% was checked by the second author who was not involved in carrying out the intervention and data collection and was blind for condition. Reliability scores were calculated using Cohen's Kappa and were: background information (1), claim (1), data (0.83), sub-arguments (0.81), warrants (0.82), counterarguments (0.92), rebuttals (0.93), qualifiers (0.93), alternative solutions (0.97) and backings (0.92).

The first author and a research assistant coded the quality aspects of the data. After the practice sessions, the first author and a research assistant independently scored a randomly selected sample of 25% of data (of both written and oral arguments). Reliability scores were calculated using Cronbach's alpha and were: overall assessment (0.87), organization of argument (0.82), sufficiency of arguments (0.89), clarity of arguments (0.81), elaboration of arguments (0.86), relevance of arguments (0.77), persuasiveness of arguments (0.79) and addressing opposing views (0.61). After resolving the disagreements the research assistant who was unaware of the data sources scored the remaining data.

5.7. Statistical analysis

As our participants came from different classes within different schools, our data were structured hierarchically. We therefore applied multilevel linear models (MLM). We used a two-level hierarchical linear model to account for the multilevel data structure with students nested within classes. We modeled the independent variables (time and condition) as fixed effects and random variations across students and classes as random effects.

To establish the effectiveness of the debate intervention, we need to take into account the combined effect of both main factors. In other words, we need to focus on the interaction of time (pre vs post) × group (intervention vs control group). For these reasons, we limit ourselves to reporting interactions.

6. Results

6.1. Structural analysis

Table 3 presents the descriptive statistics (estimated means and standard errors) for the structural analysis of written data and Table 4 for oral data.

One comment needs to be made on the basis of these statistics: the participants in both groups seem to have the tendency to heavily rely on the fundamental elements of argument structure (claim and data) in defending their positions. The other secondary argument elements were used less frequently, especially in the oral argumentation.

To answer the first research question about the impact of the debate intervention on the structural elements of the participants' argumentation, we conducted MLM analyses. To estimate the magnitude of the difference between the intervention and control groups, we used Cohen's *d* effect size (ES) when significant differences were observed. The MLM results with regard to structural analysis are presented in Table 5 for written data and in Table 6 for oral data.

Table 5 shows that the debate intervention appears to have an impact on a number of structural elements of the participants' written argument. MLM revealed that the intervention group significantly improved in terms of overall argumentation complexity ($F_{(1, 284.690)} = 5.6$; $p = .010$) with a moderate to large ES (0.59), the use of data ($F_{(1, 284.261)} = 4.7$; $p = .016$) with a moderate ES (0.51), sub-arguments ($F_{(1, 284.642)} = 9.1$; $p = .002$) with a moderate to large ES (0.72), counterarguments ($F_{(1, 281.167)} = 2.9$; $p = .045$) with a moderate ES (0.42), rebuttals ($F_{(1, 292)} = 3.1$; $p = .040$) also with a moderate ES (0.45). The students in the intervention group also showed a signicant improvement in terms of their use of qualifiers ($F_{(1, 292)} = 3.4$; $p = .034$) and backings ($F_{(1, 284.521)} = 3$; $p = .043$) with moderate effects on the two indices: 0.45 for qualifiers, 0.41 for backings.

As to oral arguments, the students in the intervention group also showed a signicant improvement in terms of overall argument complexity ($F_{(1, 286)} = 9.8$; $p = .001$) with a moderate to large ES (0.78), the number of data ($F_{(1, 278.327)} = 22.5$; $p < .001$) with a large ES (1.14), sub-arguments ($F_{(1, 278.912)} = 13$; $p < .001$) also with a large ES (0.89), warrants ($F_{(1, 279.519)} = 6.5$; $p = .006$) with a moderate to large ES (0.65), counterarguments ($F_{(1, 286)} = 4.4$; $p = .019$) with a moderate ES (0.52) and rebuttals ($F_{(1, 286)} = 7.4$; $p = .004$) with a moderate to large ES (0.69).

6.2. Argument quality analysis

Tables 7 and 8 display the descriptive statistics of argument quality analysis for both written and oral data respectively.

To answer the second research question, MLM was performed to establish whether the intervention managed to exert impact on the

Table 3
Means and standard errors of outcome variables across time and condition of written data.

Measures	Intervention group (n = 95)				Control group (n = 51)			
	Pretest		Posttest		Pretest		Posttest	
Overall argument complexity	2.63	(.14)	3.33	(.14)	2.88	(.18)	2.93	(.18)
Background information	0.10	(.04)	0.10	(.04)	0.12	(.05)	0.12	(.05)
Claim	1.00	(.01)	0.98	(.01)	1.00	(.01)	1.00	(.01)
Data	3.49	(.25)	4.20	(.25)	4.37	(.33)	4.33	(.33)
Sub-arguments	1.88	(.25)	3.05	(.25)	2.23	(.33)	2.21	(.33)
Warrants	0.13	(.06)	0.21	(.06)	0.33	(.08)	0.25	(.08)
Counterarguments	0.18	(.06)	0.34	(.06)	0.37	(.08)	0.30	(.08)
Rebuttals	0.21	(.06)	0.47	(.06)	0.35	(.09)	0.35	(.09)
Qualifiers	0.08	(.04)	0.25	(.04)	0.06	(.05)	0.06	(.05)
Alternative solutions	0.18	(.05)	0.23	(.05)	0.18	(.06)	0.29	(.06)
Backings	0.21	(.09)	0.52	(.09)	0.24	(.12)	0.30	(.12)

Table 4
Means and standard errors of outcome variables across time and condition of oral data.

Measures	Intervention group (n = 96)				Control group (n = 51)			
	Pretest		Posttest		Pretest		Posttest	
Overall argument complexity	1.79	(.09)	2.46	(.10)	1.94	(.13)	1.90	(.13)
Background information	0	(.01)	0.01	(.01)	0	(.01)	0.02	(.01)
Claim	0.99	(.01)	1.00	(.01)	1.00	(.01)	1.00	(.01)
Data	2.07	(.13)	2.99	(.13)	2.58	(.18)	2.39	(.18)
Sub-arguments	0.75	(.14)	1.95	(.14)	0.76	(.18)	0.99	(.18)
Warrants	0.04	(.04)	0.27	(.05)	0.06	(.06)	0.04	(.06)
Counterarguments	0.05	(.04)	0.24	(.04)	0.08	(.05)	0.08	(.05)
Rebuttals	0.05	(.04)	0.28	(.04)	0.12	(.06)	0.08	(.06)
Qualifiers	0	(.01)	0.02	(.01)	0	(.02)	0.02	(.02)
Alternative solutions	0.02	(.01)	0.01	(.01)	0.02	(.02)	0.02	(.02)
Backings	0.03	(.03)	0.12	(.03)	0.02	(.04)	0.02	(.04)

Table 5
Multilevel analysis results of written data.

Measures	Fixed effects				Random effects	
	Denominator df	F	P (one- tailed)	d	Variance within class	Variance between class
Overall argument complexity	284.690	5.610	.010	.59	1.20	.03
Background information	284.017	.000	1.00		.91	.00
Claim	286.532	1.104	.147		.00	4.31
Data	284.261	4.680	.016	.51	2.02	.21
Sub-arguments	284.642	9.105	.002	.72	2.57	.16
Warrants	284.216	1.606	.103		.24	.00
Counterarguments	281.167	2.893	.045	.42	.32	.00
Rebuttals	292	3.085	.040	.43	.37	.00
Qualifiers	292	3.359	.034	.45	.14	.00
Alternative solutions	292	.360	.549		.19	.00
Backings	284.521	2.993	.043	.41	.34	.02

Numerator df = 1.

quality of the arguments produced by the intervention participants in comparison to their control peers. The results are displayed in Table 9 for written data and in Table 10 for oral data.

Tables 9 and 10 reveal that the debate intervention seems to have improved the majority of the quality aspects of written as well as oral arguments. As to written arguments, MLM showed that the intervention groups significantly improved from pretest to posttest compared to the control groups in terms of organization of arguments ($F_{(1,\ 283.887)} = 11.1$; $p = .001$) with a moderate to a large ES (0.71), clarity of arguments ($F_{(1,\ 283.928)} = 4.9$; $p = .014$) with a moderate ES (0.49), elaboration of arguments ($F_{(1,\ 284.041)} = 9.2$; $p = .002$) with a moderate to large ES (0.65), relevance of arguments ($F_{(1,\ 283.953)} = 3.9$; $p = .024$) with a moderate ES (0.43) and persuasiveness of arguments ($F_{(1,\ 284.012)} = 3.3$; $p = .036$) also with a moderate ES (0.40).

With regard to oral arguments, MLM analyses revealed that the intervention group significantly outperformed the control group in

Table 6
Multilevel analysis results of oral data.

Measures	Fixed effects				Random effects	
	Denominator df	*F*	*P* (one-tailed)	*d*	Variance within class	Variance between class
Overall argument complexity	286	9.838	.001	.78	.83	.00
Background information	291	.219	.320		.01	.00
Claim	291	.537	.232		.00	.00
Data	278.327	22.492	.000	1.14	.90	.04
Sub-arguments	278.912	12.973	.000	.89	1.17	.03
Warrants	279.519	6.515	.006	.65	.15	.00
Counterarguments	286	4.364	.019	.52	.14	.00
Rebuttals	286	7.369	.004	.69	.15	.00
Qualifiers	279.997	.007	.466		.01	.00
Alternative solutions	286	.102	.375		.02	.00
Backings	279.421	1.856	.087		.07	.00

Numerator df = 1.

Table 7
Means and standard errors of outcome variables across time and condition of written data.

Measures	Intervention group ($n = 95$)		Control group ($n = 51$)	
	Pretest	Posttest	Pretest	Posttest
Overall assessment	2.67 (.13)	3.05 (.13)	2.59 (.17)	2.85 (.17)
Organization of arguments	2.32 (.17)	3.05 (.17)	2.21 (.22)	2.44 (.22)
Sufficiency of arguments	2.44 (.20)	2.89 (.20)	2.34 (.26)	2.60 (.26)
Clarity of arguments	2.34 (.17)	2.93 (.17)	2.39 (.22)	2.60 (.22)
Elaboration of arguments	2.26 (.19)	2.86 (.19)	2.42 (.25)	2.50 (.25)
Relevance of arguments	2.56 (.17)	3.10 (.17)	2.60 (.22)	2.82 (.22)
Persuasiveness of arguments	2.53 (.16)	3.04 (.16)	2.55 (.21)	2.75 (.21)
Addressing the opposing view	1.35 (.08)	1.54 (.08)	1.30 (.10)	1.63 (.10)

Table 8
Means and standard errors of outcome variables across time and condition of oral data.

Measures	Intervention group ($n = 96$)		Control group ($n = 51$)	
	Pretest	Posttest	Pretest	Posttest
Overall assessment	2.33 (.06)	2.86 (.06)	2.28 (.08)	2.31 (.08)
Organization of arguments	1.92 (.07)	2.58 (.07)	1.90 (.09)	2.00 (.09)
Sufficiency of arguments	1.89 (.09)	2.60 (.09)	1.98 (.11)	2.02 (.12)
Clarity of arguments	1.98 (.08)	2.64 (.08)	2.18 (.11)	2,08 (.11)
Elaboration of arguments	1.91 (.06)	2.55 (.06)	1.93 (.08)	2.07 (.08)
Relevance of arguments	2.21 (.08)	2.80 (.08)	2.49 (.10)	2.42 (.10)
Persuasiveness of arguments	2.19 (.08)	2.79 (.08)	2.33 (.11)	2.28 (.11)
Addressing the opposing view	1.00 (.03)	1.09 (.03)	1.06 (.04)	1.07 (.04)

Table 9
Multilevel analysis results of written data.

Measures	Fixed effects				Random effects	
	Denominator df	*F*	*P* (one- tailed)	*d*	Variance within class	Variance between class
Overall assessment	283.955	.672	.207		.38	.06
Organization of arguments	283.887	11.109	.001	.71	.37	.13
Sufficiency of arguments	284.026	1.173	.140		.52	.17
Clarity of arguments	283.928	4.870	.014	.49	.50	.11
Elaboration of arguments	284.041	9.155	.002	.65	.50	.16
Relevance of arguments	283.953	3.937	.024	.43	.42	.12
Persuasiveness of arguments	284.012	3.289	.036	.40	.48	.10
Addressing the opposing view	284.925	.905	.171		.35	.01

Numerator df = 1.

Table 10
Multilevel analysis results of oral data.

Measures	Fixed effects				Random effects	
	Denominator df	*F*	*P* (one-tailed)	*d*	Variance within class	Variance between class
Overall assessment	287	14.332	.000	.93	.29	.00
Organization of arguments	279.185	22.668	.000	1.16	.23	.01
Sufficiency of arguments	280	20.567	.000	1.08	.36	.02
Clarity of arguments	278.988	26.617	.000	1.25	.35	.02
Elaboration of arguments	280.129	13.918	.000	.91	.30	.00
Relevance of arguments	279.459	19.347	.000	1.09	.35	.01
Persuasiveness of arguments	279.283	18.595	.000	1.05	.37	.01
Addressing the opposing view	278.977	2.323	.065		.04	.00

all quality measures with a large ES except *addressing the opposing view*[6] index which approached significance: overall assessment ($F_{(1, 287)} = 14.3$; $p < .001$), organization of arguments ($F_{(1, 279.185)} = 22.7$; $p < .001$), sufficiency of arguments ($F_{(1, 280)} = 20.6$; $p < .001$), clarity of arguments ($F_{(1, 278.988)} = 26.6$; $p < .001$), elaboration of arguments ($F_{(1, 280.129)} = 13.9$; $p < .001$), relevance of arguments ($F_{(1, 279.459)} = 19.3$; $p < .001$) and persuasiveness of arguments ($F_{(1, 279.283)} = 18.6$; $p < .001$).

7. Discussion

In this study, our goal was to examine the effects of in-class debates on participants' written and oral argumentative competence. The findings, on the whole, confirmed our hypotheses and revealed that the debate intervention impacted in a positive way on a number of structural and quality aspects of the argumentative competence of the intervention participants' written and oral argumentation. Though we cannot rule out the possibility that consistent practice could have some effect on the ensuing effects, we believe that practice effect cannot merely account for the robust significant gains. This practice took place in an authentic, meaningful and interactional environment that facilitated collaborative reasoning in which the debaters appropriated argumentative strategies and awareness of what makes reasoning effective.

The developmental trends in the structural and quality aspects of written arguments were, to a large extent, similar to the development patterns in students' oral argument output. These developmental patterns further testify to the effectiveness of the intervention and suggest that there was a transfer of gains between the oral and written argumentative skills. The debate intervention facilitated a close connection between the arguments in the two modalities (i.e. writing and speaking) and hence made the transfer of gains easier. Students wrote cases to be delivered, and they assumingly carried the experience they get after participating in each debate to the next one. In this way the gains strengthened and transferred between the two modalities.

It is important to note that the intervention students did not get any instruction on the structure of argument and what promotes its quality. Rather, they seem to incrementally come to realize after participating in the intervention that convincing a critical opponent hinges on the strength and soundness of their arguments. The recurrent mission to persuade this latter has sharpened their critical thinking and provided them with insight into their reasoning capacity and grasp of what counts as good and persuasive evidence. This finding corresponds with Venville and Dawson (2010) and gives further support to Kuhn's (1991) premise that argumentation skills exist within students in a latent or embryonic state and that through participation in argumentative activities, students are able to develop the complexity of their argumentation. This suggests that developing argumentative skills require sustained practice in rich environments that entail the use of these skills (Crowell & Kuhn, 2014; Kuhn, 2018).

Another feature of debates that might have contributed to the emerged results is the presence of audience in the form of opponent peers. Berland and McNeill (2010) and Chen et al. (2016) maintain that the audience provides students with an urgent reason to develop persuasive and more complex arguments and connect oral and written arguments.

We will now further zoom in and expatiate on the findings. The intervention students significantly used more data both in their written and oral production in the posttest. The debate environment seems to foster the conviction that a considerable quantity of data is needed to defend one's point of view. Another possible explanation is that the intervention itself increased the debaters' argumentation fluency.[7] This is more apparent with oral data because during the oral task (which was semi-spontaneous) the participants had to think up arguments after only 7 minutes of preparation. Earlier research correlated data quantity with the argumentative quality of writing (e.g. Huh & Lee, 2014). The intervention students also generated significantly more sub-arguments in both modalities in the posttest. Sub-arguments represent more complex, hence more sophisticated arguments (e.g. Hoffmann, 2016; Wang, 2016). The significant increase in the use of sub-arguments to buttress main arguments (data) is an indication of the positive effect on the depth of reasoning and critical thinking of debaters.

Progress was also made in terms of the use of warrants as compared to control groups, with significance reached for the oral argumentative discourse. The presence of warrants indicates that the debate intervention infused in the students the ability to recognize the need to justify the link made between the data and the claim. Warrants, which are seldom employed in student

[6] Note that the interrater reliability for *addressing opposing views* variable is quite low, and hence its results should be interpreted with caution.
[7] We are not aware of the use of this term in literature which we would define as the ease with which a learner can think up arguments (i.e. a smooth flow of arguments).

argumentation (e.g. Crammond, 1998; Cheng, 2010), enhance the rhetoric and persuasiveness of an argument (Crammond, 1998).

The intervention also seems to have an effect on some of the secondary components of written and oral arguments though their use remained limited. The intervention students significantly raised more counterarguments and refuted them in the posttest. The employment of counterarguments and addressing them evince the ability to identify with a critical audience with an opposing perspective; this ability entails great epistemological sophistication and perspective taking (Crammond, 1998; Hays & Brandt, 1992). Responding to counterarguments is regarded as a hallmark of critical thinking (e.g. Liu & Stapleton, 2014; Nussbaum & Schraw, 2007) and as an indicator of a writer's rhetorical and reasoning competence (Cheng & Chen, 2009). It enhances the quality, persuasiveness and effectiveness of arguments (Crammond, 1998; Erduran et al., 2004; Nussbaum et al., 2005; O'Keefe, 1999).

The intervention students also significantly outperformed their control counterparts in terms of the utilization of qualifiers in their written argumentation. Crammand (1998) considers the use of qualification as an important rhetorical aspect of persuasiveness. Qualification indicates some concession on the debater' side to the audience's concerns. By placing some limit on the scope of the claim, it will sound more acceptable (Cheng & Chen, 2009). The significant increase in the use of qualification could be also seen as a sign of becoming more open-minded as the intervention students try to show more understanding for their opponents' viewpoints (see Kennedy, 2007, 2009).

Compared to the control group, the intervention group employed more backings in the posttest. The use of backings structurally extend and elaborate arguments and hence strengthen them (Cheng & Chen, 2009), and are seen as a powerful rhetorical strategy to obtain the audience's acceptance of the debater's claim (Crammond, 1998).

For a contribution to be persuasive, not only must it have arguments of good surface structure (i.e. structural sophistication), but its arguments must be cogent and have qualitative sophistication as well (e.g. Paek & Kang, 2017; Qin & Karabacak, 2010). That is, the claims need to be buttressed with sufficient, relevant, sound, clear and convincing arguments. Put differently, not only the structure, but also the content of arguments needs to be sound. In the posttest, the intervention students managed to produce qualitatively better written and oral arguments than students in the control group.

During the actual debates, especially during the rebuttal and clash stages, the students challenged each other's reasoning; they questioned the credibility and appropriateness of the advanced arguments, demanded more elaborations and justifications and critiqued reasoning inadequacies and inconsistencies. By addressing these attacks on their line of reasoning, the students seem to have fostered a critical verification of arguments and accordingly developed a sophisticated grasp of what constitutes good quality reasoning.

8. Research directions and limitations

This study has also made a case for implicit instruction in fostering L2 argumentative skills. Nonetheless, it would be interesting and worth investigating whether explicit instruction about the different structural and quality aspects of argumentation (prior to debating) would further promote the obtained gains. Moreover, it would be also interesting to investigate to what extent improvement in L2 argumentation impacts different dimensions of L2 writing and speaking skills. This area is especially relevant because finding links between high-order thinking (argumentation) and language development would have important implications for the pedagogy of L2/FL learning and teaching. Lastly, as suggested by an anonymous reviewer, future research would benefit from investigating the linguistic features, including lexical, syntactical and stylistic features that characterize strong arguments. This research is likely to afford us with tools that can inform us of how to effectively instruct argumentation.

The current study has some noteworthy limitations which open up avenues for future research. First, we have only analyzed one oral and one written task of each participant. Future research would benefit from eliciting and analyzing more data per participant. This would yield more fine-grained insights into the argumentative competence of participants. Second, although the current study provides robust empirical evidence on the effect of debate pedagogy on L2 argumentation skills, a delayed post-test could have provided additional insights into the long-term effect of the intervention. Third, as noted in the introduction of this article, one of the potential benefits of developing L2 argumentation skills is empowering L2 learners to produce good argumentative essays. Though this research suggests that the intervention students have honed their ability to produce structurally complex and qualitatively well-reasoned arguments, this does not conclusively indicate that these students will be going to correspondingly produce better argumentative essays (at least in terms of argumentation) which are longer pieces of writing and whose production is more complex and demanding. This is an interesting research direction of invaluable implications to take in future studies. Fourth, the current study tracked the effect of debate on argumentation development in an argumentative task which seems more than warranted. Nevertheless, it would be useful to examine the effect of debate pedagogy on argumentation with genres other than argumentative writing (e.g. letters and reports) to gain insights into the transferability of the effects (i.e. genre-independent effects).

9. Conclusion

Strong argumentative competence in L2 is an important educational objective; its realization can only come about if there are learning tools that guide and stimulate students to engage in cognitive processes that orient the mind towards fostering more sophisticated view of argument and evidence. The present study stakes out a case for the effectiveness of L2 in-class debates, which proved to offer fertile ground for honing L2 argumentation skills and metacognitive knowledge of argumentation. The debate intervention in the present study has improved a number of aspects of argument structure and quality of the intervention group. After the intervention, the debaters displayed a marked tendency to diversify their arguments with sophisticated structural components. They tended to bolster their positions with strong and well-grounded evidence (with more backings), highlight its link to claims (through

warrants), hedge the strength of their claims (with qualifiers), anticipate potential counterarguments and pinpoint inadequacies in them (rebuttals).

Pedagogically, this study has not only testified to the effectiveness of L2 debate as a vehicle to develop students' L2 argumentative skills, but it has also shed light on many pedagogical features that can presumably even stimulate scaffolding and refining reasoning skills beyond the L2 context since argumentation lies at the heart of education in general. These characteristics include the presence of an authentic audience (opponents, classmates and the teacher), competitive environment that prompts students to engage in rich negotiations, linkage and interplay between written and oral argumentation, and engagement in systematic and sustained practice.

We hope that the findings of this study would stimulate L2 instructors to consider employing in-class debates on a regular basis in their teaching practice. We believe that in-class debates need to become an integral part L2/FL curriculum. Debates as a teaching tool are credited and welcomed by students and hold the potential for honing other skills, including L2 language skills (e.g. el Majidi et al., 2020). So, they are worth the effort, as the potential pay-off is substantial.

Author statement

Abid el Majidi: Conceptualization; Methodology; Formal analysis; Investigation; Data curation; Writing – original draft; Writing – review & editing; Funding acquisition, **Daniel Janssen:** Conceptualization; Methodology; Validation; Writing – review & editing; Supervision; Funding acquisition, **Rick de Graaff**: Conceptualization; Methodology; Validation; Writing – review & editing; Supervision; Funding acquisition.

Acknowledgement

This work has been financed by the Dutch Research Council.

Appendix A. Framework for structural analysis of argumentation

Structural components	Examples from data
Background information	The death penalty is used for murderers, rapist and people and that commit horrible crimes.
Claim	I am against the death penalty. I have a few arguments that support my opinion.
Data 1	First of all, some innocent people might get killed.
Data backing	A study shows that over four percent of the prisoners who got convicted to death penalty in the United States were actually innocent.
Data 2	The criminal should suffer from his or her actions.
Warrant (data 2)	If you kill the criminal you make his or her life easier.
Data 3	Second, racial and socio-economic discrimination play a big role.
Subordinate argument (data 3)	It's proven that Afro-American or Latin prisoners are more likely to be convicted to death penalty than white prisoners.
Counterargument	… many people often say that the death penalty is only given to those who deserve it.
Rebuttal	but a lot of the times prosecutors are fixated on one suspect and end up pinning the crime on them.
Rebuttal backing	Many innocent people have been executed for crimes they did not commit. I am not saying every prosecutor is biased. But it is not fair to have some people receive a punishment they did not deserve just because someone believes you committed a crime.
Alternative solution	I think that living a long life behind bars is a punishment greater than death.
Alternative solution backing	Every day will be a reminder for them that they are in prison because of the bad thing that they have done. They will have to live with that guilt every day for the rest of their life.
Constraint	First of all, I agree that not every criminal should be punished but only the criminals who did something very bad like, murdering someone.

Appendix B. Scale for the assessment of the overall argumentation complexity

Complexity level	
Level 1	texts containing only a claim and data.
Level 2	texts containing a claim, data and sub-arguments and/or warrants.
Level 3	texts containing a claim, data (sub-arguments/warrants) and at least one of the following components: qualifiers, alternative solutions or backings.
Level 4	texts containing a claim, data (sub-arguments/warrants) and rebuttals.
Level 5	texts containing a claim, data (sub-arguments/warrants), rebuttals and at least one of the following components: qualifiers, alternative solutions or backings.

Appendix C. Framework for the analysis of argumentation quality

Aspects of quality		1	5
Overall assessment	The (oral) text as a whole is adequate and well-argued.		
Organization of arguments	The (oral) text is well-organized (coherent) and flows well.		
Sufficiency of arguments	The number of arguments is adequate and sufficient.		
Clarity of arguments/ comprehensibility	The (oral) text is compressible. The arguments are clearly formulated.		
Elaboration of arguments	The arguments are well-elaborated (e.g. with examples, analogies, citing authorities, etc.).		
Relevance of arguments	The arguments presented are relevant.		
persuasiveness of arguments	The arguments presented are convincing.		
Addressing the opposing view	The (oral) text addresses the opposing view(s) adequately.		

Appendix D. Examples of scoring procedures for two written texts

Example 1.

I am against legalizing abortion, because of the following reasons. **Claim**

First of all, being able to bear children is a gift of god and at the moment you decide to kill it you're deciding against the will of god. **Data 1**.

Second, abortion could lead to mental and physical pain. **Data 2**.

An example of mental pain is that a lady could get suicidal tendencies. **Subordinate argument (data 2**).

An example of the physical pain, is the pain while the operation is finding place. **Subordinate argument 2 (data 2**).

Third, while you're utilizing abortion, you are murdering a child. **Data 3**.

A baby is a human being and has also rights. **Data 4**.

Last of all, there are ladies with the wish to bear children. it is not fair against those ladies which are not even able to bear children. **Data 5**.

Argumentation structural analysis.

Overall argument complexity	Claim	Data	Sub-arguments
2	**1**	**5**	**2**

Quality of arguments analysis.

Overall assessment	Organization	Sufficiency	Comprehensibly/clarity
3	3.5	3	3
Elaboration	**Relevance**	**Persuasiveness**	**Opposing view**
3	3.5	3	1

Example 2.

I believe that abortion should stay legal. These are the arguments which I base my opinion on. **Claim**.

First and foremost, women should do whatever they want with their body. You should not be the one who makes the decisions for her. **Data 1.**

Furthermore, there are sometimes babies born which are unwanted. **Data 2.**

Mothers will not care for their baby as they should. **Subordinate argument (data 2).**

There are also personal issues involving abortion. Many mothers that don't want to raise the baby, because they want to have a successful career or finish their school. **Data 3.**

Moreover, there are several diseases which can be transferred through sexual interaction. **Data 4.**

You don't want a baby coming into this world with aids or any other disease. **Subordinate argument 2 (data 4).**

Besides, in 1970 abortion was illegal. This meant that mothers who do not wish to bring a baby into this world, had to perform unsafe abortion in secrecy. As a result, there were a lot of mothers who died after the abortion. **Data 5.**

In today's world, you will be able to go through an abortion. and you will have no after effects from the surgery. **Data 6.**

Also, mothers who are addicted to drugs, alcohol or smoking, will not prioritize the health of their baby. **Data 7.**

Lastly, people say that adoption is an option for abortion. **Counterargument.**

If there was no abortion performed, there would be no place for many babies in foster homes or abortion babies, because there isn't enough place. **Rebuttal.**

Argumentation structural analysis.

Overall argument complexity	Claim	Data	Sub-arguments	Counterargument	Rebuttal
4	**1**	**7**	**2**	**1**	**1**

Quality of arguments analysis.

Overall assessment	Organization	Sufficiency	Comprehensibly/clarity
4	4	4	4
Elaboration	**Relevance**	**Persuasiveness**	**Opposing view**
3.5	4	4	3

References

Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education, 94*(5), 765–793.

Butt, N. (2010). *Argument construction, argument evaluation, and decision-making: A content analysis of argumentation and debate textbooks*. Wayne State University. Unpublished doctoral dissertation.

Chapple, L., & Curtis, A. (2000). Content-based instruction in Hong Kong: Student responses to film. *System, 28*(3), 419–433.

Cheng, F. W. (2010). A socio-cognitive modeling approach to teaching English argumentation. *Asian ESP Journal, 6*(1), 120–146.

Cheng, F. W., & Chen, Y. M. (2009). Taiwanese argumentation skills: Contrastive rhetoric perspective. *Taiwan International ESP Journal, 1*(1), 23–50.

Chen, Y. C., Hand, B., & Park, S. (2016). Examining elementary students' development of oral and written argumentation practices through argument-based inquiry. *Science & Education, 25*(3–4), 277–320.

Connor, U. (1990). Linguistic/rhetorical measures for international persuasive student writing. *Research in the Teaching of English, 24*, 67–87.

Crammond, J. (1998). The uses of complexity of argument structures in expert and student persuasive writing. *Written Communication, 15*, 230–268.

Crowell, A., & Kuhn, D. (2014). Developing dialogic argumentation skills: A 3-year intervention study. *Journal of Cognition and Development, 15*(2), 363–381.

De Haan, P., & Van Esch, K. (2005). The development of writing in English and Spanish as foreign languages. *Assessing Writing, 10*(2), 100–116.

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.

Ferretti, R. P., MacArthur, C. A., & Dowdy, N. S. (2000). The effects of an elaborated goal on the persuasive writing of students with learning disabilities and their normal achieving peers. *Journal of Educational Psychology, 92*, 694–702.

Freeley, A. J., & Steinberg, D. L. (2005). *Argumentation and debate: Critical thinking for reasoned decision making* (11th ed.). Belmont: Wadsworth.

Gulnaz, F. (2020). Fostering Saudi EFL learners' communicative, collaborative and critical thinking skills through the technique of in-class debate. *International Journal of English Linguistics, 10*(5), 265–283.

Hays, J. N., & Brandt, K. S. (1992). Socio-cognitive development and student performance on audience- centered argumentative writing. In M. Secor, & D. Charney (Eds.), *Constructing rhetorical education* (pp. 202–229). Carbondale: Southern Illinois University Press.

Heimgärtner, R. (2013). Intercultural user interface design – culture-centered HCI design – cross- cultural user interface design: Different terminology or different approaches? In A. Marcus (Ed.), *Design, user experience, and usability. Health, learning, playing, cultural, and cross-cultural user experience* (pp. 62–71). Berlin: Springer.

Hirvela, A. (2017). Argumentation & second language writing: Are we missing the boat? *Journal of Second Language Writing, 36*, 69–74.

Hoffmann, M. H. (2016). Reflective argumentation: A cognitive function of arguing. *Argumentation, 30*(4), 365–397.

Hsu, P. S., Van Dyke, M., Chen, Y., & Smith, T. J. (2015). The effect of a graph-oriented computer- assisted project-based learning environment on argumentation skills. *Journal of Computer Assisted Learning, 31*(1), 32–58.

Hsu, P. S., Van Dyke, M., & Smith, T. J. (2017). The effect of varied gender groupings on argumentation skills among Middle school students in different cultures. *Middle Grades Review, 3*(2), 1–22.

Huh, M. H., & Lee, I. (2014). Applying Toulmin: Does peer response play a role in Korean EFL college students' revision quality? *English teaching, 69*(3), 3–23.

Iordanou, K. (2013). Developing face-to-face argumentation skills: Does arguing on the computer help? *Journal of Cognition and Development, 14*(2), 292–320.

Johnson, R. H. (2013). The role of audience in argumentation from the perspective of informal logic. *Philosophy and Rhetoric, 46*(4), 533–549.

Jonassen, D. H., & Kim, B. (2010). Arguing to learn and learning to argue: Design justifications and guidelines. *Educational Technology Research & Development, 58*, 439–457.

Karabacak, H. I., & Qin, J. (2016). Comparison of argumentative writing by American, Chinese and Turkish university students. In K. Papaja, & C. Can (Eds.), *Language in focus: Exploring the challenges and opportunities in Linguistics and English language teaching (ELT) (129–144)*. Newcastle: Cambridge scholars publishing.

Kathpalia, S. S., & See, E. K. (2016). Improving argumentation through student blogs. *System, 58*, 25–36.

Kennedy, R. (2007). In-class debates: Fertile ground for active learning and the cultivation of critical thinking and oral communication skills. *International Journal of Teaching & Learning in Higher Education, 19*(2), 183–190.

Kennedy, R. (2009). The power of in-class debates. *Active Learning in Higher Education, 10*(3), 225–236.

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

Kuhn, D. (2018). A role for reasoning in a dialogic approach to critical thinking. *Topoi, 37*(1), 121–128.

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

Liu, F., & Stapleton, P. (2014). Counterargumentation and the cultivation of critical thinking in argumentative writing: Investigating washback from a high-stakes test. *System, 45*, 117–128.

Liu, F., & Stapleton, P. (2020). Counterargumentation at the primary level: An intervention study investigating the argumentative writing of second language learners. *System, 89*, 1–15.

Long, M. H. (1996). The role of the linguistic envi-ronment in second language acquisition. In W. Ritchie, & T. Bhatia (Eds.), *Handbook of second lan-gauge acquisition* (pp. 413–468). New York: Academic Press.

el Majidi, A., de Graaff, R., & Janssen, D. (2015). Invest in what energizes students to learn: Investigating students' attitude towards debate in the foreign language classroom. *Journal of Language Teaching and Research, 6*(5), 924–932.

el Majidi, A., de Graaff, R., & Janssen, D. (2018). Perceived effect of in-class debates in second language learning. *The European Journal of Applied Linguistics and TEFL, 7*(1), 35–56.

el Majidi, A., de Graaff, R., & Janssen, D. (2020). Debate as L2 pedagogy: The effects of debating on writing development in secondary education. *The Modern Language Journal, 104*(4), 804–821.

Malloy, J. A., Tracy, K. N., Scales, R. Q., Menickelli, K., & Scales, W. D. (2020). It's not about being right: Developing argument through debate. *Journal of Literacy Research, 52*(1), 79–100.
McKenney, S., & Reeves, T. C.((2012). *Conducting educational design research*. London: Routledge.
Mercier, H. (2011). Reasoning serves argumentation in children. *Cognitive Development, 26*, 177–191.
Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34*(2), 57–74.
Michaels, S., O'Connor, C., & Resnick, L. (2008). Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. *Studies in Philosophy and Education, 27*, 283–297.
Midgette, E., Haria, P., & MacArthur, C. (2008). The effects of content and audience awareness goals for revision on the persuasive essays of fifth-and eighth-grade students. *Reading and Writing, 21*(1), 131–151.
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, E. M., Kardash, C. M., & Graham, S. E. (2005). The effects of goal instructions and text on the generation of counterarguments during writing. *Journal of Educational Psychology, 97*(2), 157–169.
Nussbaum, E. M., & Schraw, G. (2007). Promoting argument-counterargument integration in students' writing. *The Journal of Experimental Education, 76*(1), 59–92.
O'Keefe, D. J. (1999). How to handle opposing arguments in persuasive messages: A meta-analytic review of the effects of one-sided and two-sided messages. In M. E. Roloff (Ed.), *Communication yearbook* (vol. 22, pp. 209–249). Thousand Oaks, CA: Sage.
Oros, A. L. (2007). Let's debate: Active learning encourages student participation and critical thinking. *Journal of Political Science Education, 3*(3), 293–311.
Osana, H. P., & Seymour, J. R. (2004). Critical thinking in preservice teachers: A rubric for evaluating argumentation and statistical reasoning. *Educational Research and Evaluation, 10*(4–6), 473–498.
Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching, 41*(10), 994–1020.
Paek, J. K., & Kang, Y. (2017). Investigation of content features that determine Korean EFL learners' argumentative writing qualities. *English teaching, 72*(2), 101–122.
Pally, M. (1997). Critical thinking in ESL: An argument for sustained content. *Journal of Second Language Writing, 6*(3), 293–311.
Qin, J., & Karabacak, E. (2010). The analysis of Toulmin elements in Chinese EFL university argumentative writing. *System, 38*(3), 444–456.
Reznitskaya, A., Anderson, R. C., McNurlen, B., Nguyen-Jahiel, K., Archodidou, A., & Kim, S. Y. (2001). Influence of oral discussion on written argument. *Discourse Processes, 32*(2–3), 155–175.
Roy, A., & Macchiette, B. (2005). Debating the issues: A tool for augmenting critical thinking skills of marketing students. *Journal of Marketing Education, 27*(3), 264–276.
Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching, 41*(5), 513–536.
Sampson, V., & Clark, D. B. (2008). Assessment of the ways students generate argument in science education: Current perspectives and recommendations for future directions. *Science Education, 92*(3), 447–472.
Snider, A., & Schnurer, M. (2006). *Many sides: Debate across the curriculum*. New York: International Debate Education Association.
Stapleton, P., & Wu, Y. A. (2015). Assessing the quality of arguments in students' persuasive writing: A case study analyzing the relationship between surface structure and substance. *Journal of English for Academic Purposes, 17*, 12–23.
Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
Toulmin, S. (2003). *The uses of argument* (2nd ed.). Cambridge: Cambridge University Press.
Van Eemeren, F. H., de Glopper, K., Grootendorst, R., & Oostdam, R. (2015). Student performance in identifying unexpressed premises and argumentation schemes. In F. H. Van Eemeren (Ed.), *Reasonableness and Effectiveness in argumentative discourse (743–757)*. Dordrecht: Springer.
Venville, G. J., & Dawson, V. M. (2010). The impact of a classroom intervention on grade 10 students' argumentation skills, informal reasoning, and conceptual understanding of science. *Journal of Research in Science Teaching, 47*(8), 952–977.
Walker, A., & Kettler, T. (2020). Developing critical thinking skills in high ability adolescents: Effects of a debate and argument analysis curriculum. *Talent, 10*(1), 21–39.
Wang, J. (2016). On freeman's argument structure approach. *CLAR, 23–29*.
Xargia, M. (2016). A comparative analysis of Greek vs. English: Can it be argued that teenagers can argue? *Selected papers on theoretical and applied linguistics, 21*, 839–854.
Yeh, S. S. (1998). Empowering education: Teaching argumentative writing to cultural minority middle- school students. *Research in the Teaching of English, 33*, 49–83.
Zare, P., & Othman, M. (2015). Students' perceptions toward using classroom debate to develop critical thinking and oral communication ability. *Asian Social Science, 11*(9), 158–170.
Zorwick, L. W., & Wade, J. M. (2016). Enhancing civic education through the use of assigned advocacy, argumentation, and debate across the curriculum. *Communication Education, 65*(4), 434–444.