

Debate pedagogy as a conducive environment for L2 argumentative essay writing

Language Teaching Research

2026, Vol. 30(4) 1764–1788

© The Author(s) 2023



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/13621688231156998

journals.sagepub.com/home/ltr**Abid el Majidi** 

Utrecht University, Netherlands

Rick de Graaff

Utrecht University, Netherlands

Daniel Jansen

Utrecht University, Netherlands

Abstract

Debate pedagogy may hold great potential for improving second language (L2) writing skills. This study investigates this potential by examining the effects of a debate intervention on the quality of argumentative essays of Dutch secondary school students. The intervention consisted of a number of speaking and writing activities, including case-writing and note-taking. The study, which employed a pretest–posttest–delayed–posttest design with a control group, involved 135 students from eight classes at three secondary schools in the Netherlands. To measure the effects of the intervention, we analysed argumentative essays composed by the students on three occasions (pretest, posttest, and delayed posttest). The students' essays were analysed using a range of measures for fluency, syntactic and lexical complexity, accuracy, and cohesion, as well as for communicative adequacy. Multilevel analyses revealed that the intervention group made a significant improvement in a substantial number of measures in comparison to the control group. We discuss the findings in relation to key pedagogical features of the debate environment. We conclude with implications for L2 argumentative essay pedagogy.

Keywords

argumentative essays, debate, L2 writing, L2 writing pedagogy, writing development

Corresponding author:

Rick de Graaff, Utrecht University, Trans 10, Utrecht, 3512 JK, Netherlands

Email: r.degraaff@uu.nl

I Introduction

Many students, especially at secondary schools, experience difficulties with producing good argumentation essays (Wingate, 2012). This is ‘unsettling because high quality argumentative writing is expected throughout the curriculum and needed in an increasingly competitive workplace that requires advanced communication skills’ (Ferretti & Graham, 2019, p. 1345). Surprisingly, research has not paid enough attention to the teaching and learning of this important genre (Wingate, 2012), one of the most common text types that students have to master (Mei, 2006).

Argumentative essays are also important in the second language (L2) context. The ability to write effective L2 argumentative essays is considered a key indicator of L2 writing ability (Hirvela, 2017). Argumentative writing is at the heart of L2 writing assessment in many well-known standardized tests that L2 writers commonly sit. For example, one of the two writing tasks of the TOEFL (Test of English as a Foreign Language) exam requires test takers to write an essay in response to a question that asks them to express and support their opinion about a topic or issue. Similarly, the ESOL (English for Speakers of Other Languages) exam, which is common in Europe, requires examinees to write two tasks, one of which is an argumentative essay in which a range of functions are tested. These include ‘agreeing or disagreeing with a statement, giving opinions on a question, giving information or explanations, comparing and contrasting ideas and opinions, exemplifying, giving reasons and drawing conclusions’ (ESOL). Hirvela (2017) concluded that the fact that we depend highly on argumentative writing to inform us about how well students write academically underscores its importance. However, despite its manifest importance, L2 argumentative writing research has not received adequate attention (Hirvela, 2017; Pessoa et al., 2017).

Argumentative writing is a great challenge that faces L2 learners, especially secondary school students (Hirvela, 2013, 2017; Pessoa et al., 2017). Also, instructors experience the task of easing their students into argumentative writing as challenging. They do not seem to ‘be prepared to effectively scaffold argument writing’ (Pessoa et al., 2017, p. 42). They particularly lack facilitative pedagogy for L2 argumentative writing (Hirvela, 2013, 2017). Yet, Hirvela (2013) remarked that once equipped with the right pedagogy, instructors will be able to make L2 argumentative writing accessible and manageable for L2 learners.

A number of studies have explored different pedagogical approaches that can scaffold students’ ability to produce good and well-reasoned argumentative essays (e.g. Jin et al., 2020; Matos, 2021). However, these studies were predominantly conducted in the first language (L1) context (Hirvela, 2017; Huang & Jun Zhang, 2020). L1 writing pedagogies do not always fit the L2 context (Huang & Jun Zhang, 2020). L2 argumentative writing is challenging because learners have to deal with both language and argumentation (Jin et al., 2020). One promising pedagogical approach that seems to cater to both language and argumentation in the L2 context is debate pedagogy.

Debate pedagogy – which involves debate-related activities, including mainly reading articles, writing cases,¹ and involvement in actual debates – is believed to hold promise as a conducive mechanism for advancing many textual and content features related to argumentative essays. It enables learners to attend to, and get involved in, language

processes that facilitate L2 writing development (El Majidi et al., 2020). However, the existing evidence about the conduciveness of debate pedagogy to improving argumentative essay writing is largely anecdotal. We are not aware of any experimental study into the effects of debating on different dimensions of argumentative essay writing in an L2 context. This study aims to fill this gap and hence contribute to lacking empirically-based pedagogical knowledge about how to improve L2 argumentative writing (see Hirvela, 2017). The primary objective of this study was, therefore, to provide empirical evidence about the extent to which debate pedagogy improves different dimensions of L2 argumentative essays. To this end, we analysed linguistic features relating to accuracy, fluency, syntactic and lexical complexity, and cohesion, as well as communicative adequacy.

II Theoretical underpinning of debate as an effective L2 writing pedagogy

In this section, we discuss a number of relevant theoretical and pedagogical perspectives that underlie our hypothesis that debating may be an effective pedagogical framework for L2 argumentative writing development.

Manchón (2011) discussed three writing perspectives, all of which are relevant to L2 writing instruction:

- ‘learning-to-write’ (LW);
- ‘writing-to-learn-content’ (WLC); and
- ‘writing-to-learn-language’ (WLL).

LW is a traditional perspective that sees writing as an end in itself. In contrast, the WLC perspective conceptualizes writing as a vehicle for learning disciplinary subject matter in the content areas, and WLL orientation regards writing as a means of promoting language learning mainly through raising L2 learners’ awareness of problematic linguistic areas in their output.

The debate pedagogical environment facilitates the coexistence of the three writing perspectives, and this coexistence can lead to substantial benefits for L2 writing development (see Ortega, 2011). First, writing in the debate context seems to resonate with Hyland’s (2011) view of a successful LW implementation, in which a text is conceptualized as social and reader-oriented discourse that imparts the writer’s thoughts and perspectives. Second, with debate pedagogy, students engage in writing (i.e. case writing) with a communicative purpose in mind: defending their standpoints and criticizing those of their opponents. Here, writing serves the function of synthesizing, analysing, and organizing arguments. This orientation intersects with the WLC perspective. Lastly, the competitive debate environment stimulates the use of accurate and sophisticated language to confer cogency on the adduced arguments. In other words, the debate environment provides students with a purposeful and meaningful context in which accurate and sophisticated language serves a relevant function, namely increasing the persuasiveness of the marshaled arguments and hence outshining critical opponents. This pursuit encourages

debaters to take the provided feedback on language use more seriously (El Majidi et al., 2020). This self-conscious use of language seems to correspond to the WLL orientation.

Another relevant perspective that gave rise to our hypotheses stems from the work of Merrill Swain. Swain's (1993) Output Hypothesis suggests that output prompts learners to process language more deeply and effectively than do reading and/or listening alone. As a result of inherent features of written output, including the availability of time and the visibility of the written text, the act of writing can be advantageous to language development in multiple ways (Manchón & Williams, 2016). Because it is offline, textual production offers learners more space to reflect upon their composition, notice gaps in their L2, and seek ways to remedy them (Manchón & Williams, 2016). Aside from the oral output prompted during actual debates, debate pedagogy entails the production of written output as well, as debaters summarize preparatory articles, make notes, and write cases.

Another factor, which predicates the assumption that debating can lead to improvement in L2 argumentative writing, has to do with how writing is perceived in the debate context. Writing in this context is viewed as a socially oriented activity that involves an authentic audience: Learners prepare cases for a specific audience (teacher and classmates, especially opponents). Research has demonstrated that an audience offers students extra impetus to develop rich, complex, and persuasive reasoning (Chen et al., 2016; Cho & Choi, 2018). Interestingly, audience awareness is also found to positively impact different aspects of written texts (e.g. Berland & McNeill, 2010; Cho & Choi, 2018; Turgut, 2009; Yasuda, 2019).

Debate pedagogy creates an instructional atmosphere that leverages the potential of talk (speaking) in the service of written language. The debate environment involves learners in rich and multidimensional interactions that benefit the writing modality in multiple ways. Weissberg (2006) regarded writing classrooms that integrate dialogic social interactions 'as a place where oral language is recognized as a developmental springboard into writing for L2 . . . and where a multitude of opportunities exist – some planned, some fortuitous – for dialogue to serve the purposes of writing instruction' (p. 26). Equally, Kuhn et al. (2016) regarded dialogic social argumentation 'as a path to the development of individual argumentative thinking and writing' (p. 9). In a recent empirical study, we made a case for the effects of debate pedagogy on speaking skills (El Majidi et al., 2021a). We surmise that the improvement of speaking skills in debates is likely to affect writing skills as well. Those findings, therefore, further reinforce our hypothesis that debate pedagogy holds great potential for catering to writing skills in general and argumentative writing in particular.

A number of studies have indicated that debate pedagogy can lead to improvement in reasoning/argumentative skills (Oros, 2007; Zorwick & Wade, 2016). Debate confronts debaters with plenty of conflicting facts, assumptions, and perspectives that demand the use of higher-level reasoning strategies. Involvement in debate therefore encourages students to critically analyse the opposing side's reasoning and evidence, and to identify inconsistencies and inadequacies in their line of reasoning. In addition, debate pedagogy integrates two argumentation perspectives: learning-to-argue and arguing-to-learn (see Hirvela, 2017), and this integration benefits argumentative writing in several ways (Zou et al., 2021).² Through active involvement in debates, debaters foster different strategies and dimensions of argumentation even when they do not receive instruction about

argumentation (El Majidi et al., 2021b). This promotes the learning-to-argue orientation. The arguing-to-learn perspective construes argumentation as a tool through which issues and disputes can be resolved (Jonassen & Kim, 2010). This conceptualization of argumentation is prominent in debate as argumentation is seen as a means by which debaters defend their standpoints. Also, defending one's position effectively entails developing a deeper understanding of content. In short, debate enables debaters to develop knowledge of both argumentation and content simultaneously.

Research has provided empirical evidence that debate pedagogy can hone L2 argumentation skills. In a recent study, we showed that debate pedagogy provides fertile ground for developing L2 argumentation skills and metacognitive knowledge of argumentation (El Majidi et al., 2021b). The debaters in that experimental study displayed a marked tendency to diversify their arguments with sophisticated structural components. They also tended to support their standpoints with cogent and well-reasoned evidence.

Lastly, learners' positive attitude towards debating is another factor that may boost the learning process in the debate context (e.g. Lustigova, 2011). When motivated, 'the student will engage in the (often difficult) task of writing and develop stronger skills as a writer' (Wright et al., 2021, p. 607).

III Debate-argumentative writing research

Very few studies (either in the L1 or L2 context) have explored the effects of in-class debates on writing performance in general and on argumentative writing in particular. These few studies, which drew on anecdotal evidence and mainly involved students in higher education, have suggested that the debating environment can be advantageous to writing performance (e.g. Lustigova, 2011). There are, however, two experimental studies that are noteworthy here. The first was conducted by Kimura (1998), who compared a group of university students that debated before writing with a group that only focused on writing (without debating). To gauge the effectiveness of the intervention, Kimura used expository essays, which she analytically assessed on the basis of three components: argumentation, organization, and communicative quality, using a 1–9 scale. The results showed that the debating group produced better expository essays. Kimura concluded that 'debating was a very effective way to improve the students' writing' (p. 28). However, this study, though it reveals relevant insights, has not specified which specific linguistic features of expository essays were affected by debating. In another study, El Majidi et al. (2020) examined the effects of debating on free-opinion tasks. This study showed that participation in debates resulted in improving a number of aspects of writing performance, including fluency, complexity, and accuracy. Yet, the texts produced in this study were relatively short, and hence these findings cannot be assuredly taken to postulate that debaters can compose qualitatively good longer structured argumentative texts (i.e. argumentative essays).

IV This study

In short, several theoretical and pedagogical perspectives suggest that debating can be conducive to improving argumentative essay writing. However, there is little empirical

research that has studied the extent to which this environment can affect different areas of argumentative essay writing. The present study is an attempt to fill part of this research gap and equally respond to the call (e.g. Hirvela, 2017) to identify pedagogical tools that promote L2 argumentative writing. We will track the effects of debate pedagogy on argumentative essays across the dimensions of complexity, accuracy, fluency, adequacy, and cohesion (CAFAC). CAFAC measures have been identified as key indicators for L2 writing performance (including essays) and development (e.g. Crossley, 2020; Kuiken & Vedder, 2017). This study was guided by the following research question:

- Research question: What are the effects of debate pedagogy on different aspects of L2 argumentative essay writing, including fluency, syntactic and lexical complexity, accuracy, cohesion, and communicative adequacy of Dutch secondary students?

Given the theoretical and pedagogical considerations (discussed above) and previous research (e.g. Kimura, 1998), we hypothesized that students engaged in debate pedagogy would produce linguistically and communicatively better argumentative essays than students in the control group.

V Method

I Participants

The study comprised eight intact classes at three secondary schools in the Netherlands ($N=135$). Four classes were in the fifth year of higher general secondary education (*havo 5* in Dutch) ($n=65$), and four classes were in their fourth year of preuniversity secondary education (*vwo 4* in Dutch) ($n=70$). Four classes served as the intervention group ($n=67$) and four as the control group ($n=68$). To control for the idiosyncratic effects associated with instructors, the intervention and control classes shared the same instructors. The participants consisted of 77 females and 58 males, ranging in age from 15 to 18 years. The English proficiency level of all classes (including writing) roughly spanned the B1 and B2 levels as estimated by their teachers and related to Dutch educational guidelines. B1 and B2 levels are the third and fourth levels, respectively, of English in the Common European Framework of Reference for Languages (CEFR); these levels are comparable to the intermediate and upper intermediate levels.

2 Intervention

We previously conducted an educational design research (see McKenney & Reeves, 2012) on the basis of which the current debate intervention was engineered. The intervention students participated in 10 debates (one per week) and were informed at least one week in advance of the debate topic to make the necessary preparation. The debate topics (e.g. the right to bear arms) were selected with their consent. Prior to the start of debates, we familiarized the students with the mechanics of the debate task (see Table 1). We discussed how debates should run and what students have to do in each stage. During this

Table 1. Main writing activities performed during the intervention.

Intervention group	Control group
<p><i>Pre-debate stage:</i></p> <ul style="list-style-type: none"> – Reading two articles and summarizing them. The instructor provided one article and the students had to find another one – Writing cases of 80 words <p><i>During-debate stage:</i></p> <ul style="list-style-type: none"> – Noting down the arguments of the opponents, as debaters had to rebut them during the rebuttal and clash stages – Noting down new words and mistakes from classmates <p><i>Post-debate stage:</i></p> <ul style="list-style-type: none"> – Processing the feedback provided by the instructor on written cases 	<ul style="list-style-type: none"> – Reading and summarizing newspaper articles (e.g. from <i>The Guardian</i>), covering current issues, including argumentative articles, addressing, for example, political and changes in policy issues – Mainly writing argumentative essays and letters, including complaint letters in which students had to express dissatisfaction with a particular service and accordingly convince the addressed company to provide a refund

discussion – which was not extensive (about 45 minutes) since many of the debating skills are natural communication practices that are used daily (Snider & Schnurer, 2006) – we asked our students to pay attention to the evidence they use to support their claims, but we did not provide further instruction on argumentation. We did not, for example, discuss the components of arguments or types of argumentation, nor did we talk about the different aspects of argument quality being measured in this study (see communicative adequacy).

All debates (in the during-debate stage) had three phases: constructive speech, rebuttal, and clash. We employed two debate formats: debating in a group of four debaters (two students in favor and two against) and a one-to-one debating format. Each debate consisted of three stages: pre-debate, during-debate, and post-debate. Table 1 presents the writing activities performed in each debate session, in addition to the writing tasks undertaken in parallel by the control group.

It is important to note that the intervention students also wrote essays and letters, but fewer than the control students. Two control classes wrote on average one more essay and letter (including at least 200 words each) than their corresponding intervention peers. As to the number of classroom sessions, all the intervention classes and their corresponding control ones followed the same number of sessions. The intervention and control groups received regular instruction comprising activities covering the four language skills, and they both received feedback on their produced texts. The control group used the same course materials as the intervention group but did not participate in any debate-related activities. In place of the debate intervention (once a week), learners in the control group received regular lessons in which their language skills were further practiced. In short, the control and intervention groups received the same number of sessions per week, were instructed by the same teachers, and received the same teaching material; the only difference is that while the intervention group debated once a week, the control group practiced more with language skills, including writing.

3 Procedure

To measure the effects of the debate intervention on argumentative essays, we adopted a pretest–posttest–delayed posttest control group design. We compared three essays of comparable familiarity and difficulty previously administered to similar classes: before the intervention (pretest), after the intervention (posttest), and approximately three months later (delayed posttest). We selected three topics: (1) violent computer games cause behavior problems; (2) the internet does more harm than good; and (3) the legal age to attain a driver's license should be raised to 21 years. These topics are accessible and of interest to students in this age group, and we expected our students to have ample exposure to the background information related to these topics. The topics were identical for all participating classes and were counterbalanced to avoid any potential topic effect. The participants received instructions about the overall structure of essay writing and were required to complete the argumentative essays using at least 200 words within 50 minutes. The conditions around the completion of the essays were the same for both the intervention and control groups. Due to practical constraints, we only managed to collect essays for two classes on the delayed posttest occasion; these two classes both belong to the intervention group ($n=36$).³ Ethical approval for the research was granted by the researchers' institution, and consent was also obtained from the students' parents.

4 Measures

We included a large range of measures in this study to capture a comprehensive picture of the developmental patterns induced by the intervention. We followed two principles in choosing our measures: (1) selecting common measures used in comparable studies to ensure comparison with previous research; and (2) each measure should address a specific facet of the construct in question. We assessed the essays for quality of linguistic production measured with metrics tapping into fluency, syntactic and lexical complexity, accuracy, and cohesion dimensions – which reveal key insights into writing quality and development (e.g. Crossley, 2020; Michel, 2017) – and for the quality of communicative adequacy – which reflects fundamental insights into the content of the written production (e.g. Kuiken & Vedder, 2017). The measures were a mixture of automatically coded features and of measures that required hand coding.

a Fluency. Fluency was measured in terms of text length (i.e. total number of words produced). This is the most common metric for measuring written fluency (Plakans et al., 2016).

b Syntactic complexity. To measure syntactic complexity, we used three indices recommended by Norris and Ortega (2009) that span three dimensions of this construct:

- global complexity: number of words per T-unit (MLT);
- complexity by subordination: mean number of clauses per T-unit (C/T);
- clausal/phrasal complexity: mean length of clauses (MLC).

In this study, all indices of syntactic complexity were measured by the automatic L2 syntactic complexity analyser (Lu, 2010), which was specifically developed to parse L2 written data.

c Accuracy. To measure accuracy, we first segmented the essays into clauses following Miller's (2008) guidelines. Then, we calculated the ratio of error-free clauses, which is widely recognized as a reliable global measure for tracking changes in accuracy (e.g. Tavakoli & Skehan, 2005). In addition, we calculated the number of errors per 100 words for different linguistic categories (for the operationalization and examples of the first three measures, see Yoon & Polio, 2017). Spelling and punctuation errors were ignored unless a misspelled word resulted in an actual English word (Ferris & Roberts, 2001). We computed the following indices:

- error-free clauses (EFC);
- lexical errors per 100 words;
- syntactic errors per 100 words;
- morphological errors per 100 words;
- prepositional errors per 100 words.

d Lexical complexity. We used two measures of lexical sophistication and one measure of lexical diversity obtained from the computational tool Coh-Metrix (McNamara et al., 2010):

- measure of textual lexical diversity (MTLD);
- average word length;
- word frequency.

We used the MTLD index to measure lexical diversity, since it is less affected by text length and allows for comparison between texts of different lengths (McNamara et al., 2010). To measure lexical sophistication, we used two indices that are seen as reliable predictors of sophisticated vocabulary: average word length, which reflects mean word length, with longer words indicating more sophistication; and word frequency index, which calculates the mean logarithmic frequency for all words. A lower word frequency (WF) indicates higher sophistication.

e Cohesion. To track the effect of the intervention on cohesion, we adopted Hyland's (2005) framework for interactive meta-discourse. Hyland's analytic framework, which is widely used in the field of L2 academic writing research (Takač & Ivezić, 2019), offers a fine-grained analysis of cohesive devices, since each marker is assessed in its own right. Research has shown that the frequency and diversity of meta-discourse markers significantly reflect the quality of argumentative texts (Qin & Uccelli, 2016).

Following Hyland's procedures, we coded the essays for four types of organizational markers, in addition to their diversity of type (see, for example, Dobbs, 2014; Qin & Uccelli, 2016) and token:

- frame markers: markers that mark the sequence of arguments (e.g. *firstly*);
- code glosses markers: markers that introduce an example or paraphrase (e.g. *for example*);
- transition markers: markers that mark additive, adversative, or causal relations between clauses and paragraphs (e.g. *besides, although, because*); temporal markers and the coordinating conjunction *and* were excluded since they are less associated with quality (Dobbs, 2014);
- conclusion markers: markers that introduce a summary or conclusion (e.g. *all in all*);
- markers diversity token: diversity of markers in terms of token;
- markers diversity type: diversity of markers in terms of type.

f Communicative adequacy. To obtain a full and reliable picture of performance, many scholars have stressed that it is imperative to take the communicative dimension of L2 production into consideration (also known as ‘functional adequacy’) as an essential component of L2 proficiency (e.g. De Jong et al., 2012; Kuiken & Vedder, 2017, 2022).

Communicative adequacy is a task-related construct (Kuiken & Vedder, 2017, 2022) and can be measured in multiple ways, for example, by means of qualitative ratings (Pallotti, 2009). In this study, we used a five-point Likert rating scale that was developed in the study of El Majidi et al. (2021b). This rating scale covers the four dimensions of communicative adequacy (i.e. task requirement, content, comprehensibility, and coherence/cohesion) proposed by Kuiken and Vedder (2017, 2022). It assesses different dimensions of argumentation – which is at the heart of the communicative success of the present task – in addition to features that ensure the presentation of an organized and structured essay. These dimensions include:

- overall assessment, i.e. the extent to which the essay as a whole is adequate and well argued;
- organization, i.e. the extent to which the arguments are well organized;
- sufficiency, i.e. the extent to which the number of arguments is sufficient and adequate;
- comprehensibility/clarity, i.e. the extent to which the arguments are clearly formulated;
- elaboration, i.e. the extent to which the arguments are well elaborated, e.g. with examples, analogies, citing authorities, etc.;
- relevance, i.e. the extent to which the arguments presented are relevant;
- persuasiveness of arguments, i.e. the extent to which the arguments presented are convincing;
- addressing the opposing view, i.e. the extent to which the opposing view is adequately addressed.

5 Interrater reliability

The hand-coded measures (i.e. accuracy, cohesion, and adequacy measures) were initially coded by the first author. To measure interrater reliability for these measures, a

randomly selected sample of 20% of the total data was checked by a trained research assistant, who coded similar data in previous studies and was masked to condition. Cohen's kappa was high for both accuracy and cohesion measures. A minimum interrater reliability coefficient of .95 was achieved for each measure. Interrater reliability for adequacy measures (scale-based), which was calculated by Cronbach's alpha, was also high; all measures exceeded .89.

6 Statistical analysis

As our participants came from different classes within different schools, our data were structured hierarchically. We therefore applied multilevel linear model analyses (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. Thus, we merely focus on, and report, the interaction of time (pretest vs. posttest) \times group (intervention vs. control group). Prior to performing statistical analyses, we checked the prerequisite assumption that the different residual scores are normally distributed through a visual inspection of the histograms of each residual, which is the standard procedure in multilevel modeling. No notable deviations were visible.

VI Results

Table 2 presents the descriptive statistics (estimated means and standard errors) for the scores obtained (at the three test time points) for each group on fluency, syntactic and lexical complexity, accuracy, and cohesion measures. Table 3 illustrates these statistics for the communicative adequacy measures. Results of both Tables 2 and 3 show that the groups' means increased on most measures over the intervention period and three months later.

To address the research question of the study, we conducted MLM analyses. To estimate the magnitude of the debate intervention effect, we compared the impact of the intervention to the total variance (Cohen's d) when significant differences were observed. The MLM results for fluency, syntactic and lexical complexity, accuracy, and cohesion measures are presented in Table 4 and in Table 5 for communicative adequacy.

Table 4 shows a significant increase for the intervention group in terms of fluency, $F(1, 284.992) = 12.2, p = .001$, with a moderate to large effect size (ES) (see Cohen, 1988) ($d = 0.76$). While the control group produced on average two more words in the posttest than in the pretest, the intervention group produced 55 more words in the posttest and 34 more words in the delayed posttest.

As for syntactic complexity, the intervention group comparatively showed a significant improvement in terms of global complexity (MLT), $F(1, 285.417) = 5.6, p = .010$, with a moderate ES ($d = 0.58$) and in terms of subordination (C/T), $F(1, 285.871) = 4.2, p = .020$, again with a moderate ES ($d = 0.52$). While C/T slightly decreased in the delayed posttest, MLT gains were further reinforced.

Table 2. Means and standard errors across time and condition of fluency, syntactic and lexical complexity, accuracy, and cohesion.

Measures	Index	Intervention group (n = 67)			Control group (n = 68)		
		Pretest	Posttest	Delayed posttest	Pretest	Posttest	Posttest
Fluency	Number of words	252.0 (18.55)	307.59 (18.73)	286.24 (20.43)	251.32 (18.50)	253.72 (18.55)	
Syntactic complexity	MLT	12.43 (.52)	13.71 (.53)	14.46 (.63)	13.72 (.52)	13.48 (.52)	
	MLC	7.41 (.17)	7.80 (.17)	8.40 (.22)	7.46 (.17)	7.74 (.17)	
Lexical complexity	C/T	1.70 (.06)	1.77 (.07)	1.74 (.08)	1.87 (.06)	1.76 (.06)	
	MLTD	60.63 (3.00)	64.43 (3.06)	65.33 (3.57)	58.92 (2.99)	58.91 (3.01)	
Accuracy	Word frequency	3.13 (.02)	3.15 (.02)	3.17 (.02)	3.17 (.02)	3.18 (.02)	
	Word length	4.34 (.04)	4.34 (.04)	4.35 (.05)	4.35 (.04)	4.34 (.04)	
Cohesion	Error-free clauses	0.67 (.02)	0.77 (.02)	0.78 (.03)	0.69 (.02)	0.73 (.02)	
	Lexical errors	1.36 (.11)	0.79 (.12)	0.73 (.15)	1.34 (.11)	1.20 (.12)	
Syntactic errors	Syntactic errors	0.94 (.09)	0.61 (.09)	0.39 (.12)	0.86 (.09)	0.93 (.09)	
	Morphological errors	3.63 (.35)	2.50 (.36)	2.28 (.44)	3.54 (.35)	2.35 (.35)	
Preposition errors	Preposition errors	0.46 (.07)	0.33 (.07)	0.39 (.09)	0.49 (.07)	0.44 (.07)	
	Transition markers	6.92 (.46)	9.14 (.47)	8.21 (.61)	6.84 (.45)	6.70 (.46)	
Frame markers	Frame markers	0.77 (.21)	1.86 (.21)	2.68 (.27)	0.68 (.21)	1.11 (.21)	
	Gloss markers	1.04 (.20)	1.07 (.21)	1.02 (.27)	1.29 (.20)	0.92 (.21)	
Conclusion markers	Conclusion markers	0.20 (.06)	0.59 (.06)	0.75 (.08)	0.24 (.06)	0.50 (.06)	
	Diversity type	2.15 (.11)	2.80 (.11)	3.25 (.15)	2.32 (.11)	2.59 (.11)	
Diversity token	Diversity token	5.34 (.32)	8.18 (.33)	9.01 (.43)	5.45 (.32)	6.15 (.32)	

Notes. MLT = number of words per T-unit. MLC = mean length of clauses. C/T = mean number of clauses per T-unit. MLTD = measure of textual lexical diversity.

Table 3. Means and standard errors across time and condition of communicative adequacy.

Measures	Intervention group (n=67)			Control group (n=68)	
	Pretest	Posttest	Delayed posttest	Pretest	Posttest
Overall assessment	2.62 (.14)	3.10 (.14)	3.17 (.17)	2.64 (.14)	2.72 (.14)
Organization of arguments	2.49 (.12)	3.16 (.12)	3.22 (.15)	2.53 (.12)	2.67 (.12)
Sufficiency of arguments	2.83 (.17)	3.26 (.17)	3.35 (.20)	2.78 (.17)	2.80 (.17)
Comprehensibility of arguments	2.55 (.12)	2.95 (.12)	3.04 (.15)	2.56 (.12)	2.65 (.12)
Elaboration of arguments	2.79 (.16)	3.17 (.16)	3.33 (.19)	2.67 (.16)	2.73 (.16)
Relevance of arguments	2.67 (.14)	3.15 (.14)	3.21 (.16)	2.66 (.14)	2.72 (.14)
Persuasiveness of arguments	2.51 (.13)	3.07 (.13)	3.15 (.16)	2.56 (.13)	2.60 (.13)
Addressing the opposing view	2.41 (.17)	2.99 (.17)	3.02 (.19)	2.35 (.17)	2.47 (.17)

Table 4. Multilevel analysis results for fluency, syntactic and lexical complexity, accuracy, and cohesion.

Measures	Index	Fixed effects				Random effects	
		Denominator df	F	P (one-tailed)	d	Variance within class	Variance between classes
Fluency	Number of words	284.992	12.207	.001	0.76	3660.57	1144.15
Syntactic complexity	MLT	285.417	5.587	.010	0.58	6.82	.67
	MLC	287.645	0.152	.348		1.22	.04
	C/T	285.871	4.249	.020	0.52	.11	.01
Lexical complexity	MLTD	286.172	1.175	.140		197.15	23.67
	Word frequency	287.645	.000	.500		.01	.00
	Word length	286.265	.096	.378		.05	.00
Accuracy	Error-free clauses	286.936	3.374	.034	0.46	.02	.00
	Lexical errors	283.621	3.708	.028	0.49	.76	.01
	Syntactic errors	285.759	5.253	.012	0.71	.50	.00
	Morphological errors	285.796	.012	.457		4.23	.23
	Preposition errors	285.709	.377	.270		.22	.01
Cohesion	Transition markers	285.228	8.330	.002	0.72	10.66	.17
	Frame markers	286.879	3.982	.024	0.43	1.76	.06
	Gloss markers	285.705	1.439	.116		1.70	.06
	Conclusion markers	285.737	1.301	.128		.22	.00
	Diversity type	293	2.860	.046	0.43	.79	.00
	Diversity token	284.645	12.572	.000	0.85	5.87	.05

Notes. MLT=number of words per T-unit. MLC=mean length of clauses. C/T=mean number of clauses per T-unit. MLTD=measure of textual lexical diversity. Numerator $df=1$.

Table 5. Multilevel analysis results for communicative adequacy.

Measures	Fixed effects				Random effects	
	Denominator <i>df</i>	<i>F</i>	<i>P</i> (one- tailed)	<i>d</i>	Variance within class	Variance between class
Overall assessment	285.908	6.072	.007	0.58	.43	.05
Organization of arguments	285.514	13.074	.000	0.86	.35	.04
Sufficiency of arguments	285.254	5.957	.008	0.56	.45	.09
Comprehensibility of arguments	286.404	3.484	.032	0.46	.44	.03
Elaboration of arguments	285.507	3.328	.035	0.43	.48	.07
Relevance of arguments	285.973	6.303	.007	0.60	.43	.05
Persuasiveness of arguments	286.512	8.989	.002	0.71	.50	.03
Addressing the opposing view	285.525	7.528	.003	0.63	.45	.08

Note. Numerator *df*= 1.

No significant differences were found for lexical complexity measures, in spite of modest observable differences in means for the MTLT measure in favor of the intervention group.

With regard to accuracy, the intervention group significantly outperformed the control group in terms of three measures: error-free clauses, $F(1, 286.936)=3.4, p=.034$, with a moderate ES ($d=0.46$); lexical errors, $F(1, 283.621)=3.7, p=.028$, again with a moderate ES ($d=0.49$); and syntactic errors, $F(1, 285.759)=5.3, p=.012$, with a moderate to large ES ($d=0.71$). The obtained gains were further improved at delayed posttest.

With respect to cohesion, the intervention group comparatively displayed an improvement across all measures of cohesion, with four measures reaching statistical significance: transition markers, $F(1, 285.228)=8.3, p=.002$, with a moderate to large ES ($d=0.72$); frame markers, $F(1, 286.879)=4.0, p=.024$, with a moderate ES ($d=0.43$); diversity of type, $F(1, 293)=2.9, p=.046$, again with a moderate ES ($d=0.43$); and diversity of token, $F(1, 284.645)=12.6, p<.001$, with a large ES ($d=0.85$). Importantly, the intervention group further improved most of the scores of cohesion measures at delayed posttest.

To examine the extent to which the observed (significant) effects in the posttest were retained in the delayed posttest (i.e. long-term effects), we performed contrast tests (contrasting the posttest and delayed posttest scores). These tests showed that the difference between the second and third measurement occasion was significant for frame markers ($t(293)=-2.85, p=.005$) and diversity of type ($t(293)=-2.39, p=.018$) measures. In these measures, the intervention students further significantly improved the progress demonstrated in the posttest. As for other measures, the difference was insignificant, meaning that the observed effects at posttest sustained over time. As Table 5 shows, the intervention group outperformed the control group in all dimensions of communicative adequacy, with moderate to large ES, ranging from $d=0.43$ to $d=0.86$.

To examine the durability of the observed effects of communicative adequacy in the posttest, we performed contrast tests. The contrast tests revealed that the difference between the posttest and delayed posttest was statistically insignificant. This means that the intervention group maintained the improvement of the posttest in the delayed posttest.⁴

VII Discussion

The primary aim of this study was to assess the impact of debate pedagogy on different dimensions of argumentative essay writing. We charted progress on six main textual constructs (i.e. CAFAC). Our overall results show that the intervention group significantly outperformed the control group in most of the examined dimensions of essay argumentative writing. After the intervention, the debate group comparatively produced argumentative essays that were longer, exhibited more complex syntactic structures, and were couched in a more diverse, coherent, and persuasive language. These results are consistent with our hypothesis and are congruent with previous research that showed that debate pedagogy empowers students to construct good essays both in L1 (Mirra et al., 2016) and L2 (Kimura, 1998).

Ostensibly, it may seem that the observed effects of debate pedagogy are merely practice effects because the intervention group produced, on average, more writing output than the control group. However, the results cannot be exclusively ascribed to such practice effects. Only practice in writing without, for example, dialogic argumentation would not 'yield the same benefits' (Kuhn et al., 2016, p. 136). There are several explanations for the observed improvements. The intervention weaves together several instructional practices and factors that have proven to be effective in earlier research for L2 writing.

Debate pedagogy engages students in a meaningful learning experience that facilitates a meaningful link between input and output. This environment benefits writing in many ways. Writing in the debate environment is a goal-oriented activity that gives debaters the sense that they are important agents that create meaning that affects people through the arguments they present. This context requires debaters to ponder over the substance of their contributions (i.e. arguments) and on how to formulate them (i.e. language). As such, it stimulates them to reflect on and critique their learning by reviewing their performance. In such an environment, the very act of writing promotes linguistic and rhetorical processing that potentially leads to argumentative language development. In a word, writing in the debate context provides a platform for learning argumentative writing (i.e. writing-to-learn) (Manchón, 2011).

Writing in the debate context is a social activity that serves a purposeful goal. The debaters seemed to find purpose in their writing; they 'are no longer writing to or for a teacher, seeking to produce what they think the teacher is looking for. Instead, they engage with peers' (Kuhn et al., 2016, p. 86) and write 'to be heard and to communicate their ideas' (Dickson, 2004, p. 35). In a context like this, students become motivated and are stimulated to produce accurate, sophisticated, and persuasive output, knowing that their work will be delivered to a group of critical students, and that good quality of output will help them to outshine these opponents. This seemingly intensifies the functionality

of sophisticated output. This means that ‘debaters must use clear, concise, powerful language to defend their positions’ (Freeley & Steinberg, 2005, p. 34). In other words, in such a context, debaters are ‘being pushed toward the delivery of a message that is not only conveyed, but that is conveyed precisely, coherently, and appropriately’ (Swain, 2005, p. 473). Research has shown that audience awareness is advantageous to L2 writing and to language development in general. For example, Chen (2019) pointed out that when students present or write something for a real audience, they become motivated to reflect upon their production. Similarly, Sasaki et al. (2020) found that writing tasks that involve a real audience can improve L2 writing fluency and motivation for writing. In short, as Chen and Brown (2012) maintained, the presence of an authentic audience in a challenging environment increases L2 learners’ motivation and effort ‘to carefully choose appropriate words or phrases to express their ideas and deliberately use sophisticated structures’ (p. 446).

In addition, debate pedagogy facilitates a recursive writing process and a synergetic connection between language skills. Prior to each debate, the intervention group read argumentative texts and subsequently composed (argumentative) cases. Hirvela (2016) pointed out that this connection leads to ‘reflecting something from the original sources while at the same time shaping that content in a new way relative to the ultimate purpose served by their writing, such as using sources to generate an argumentative essay or a research paper’ (p. 44). Mirra et al. (2016) asserted that reading in the context of debate is purposeful because debaters are provided with ‘an authentic reason to read’ (p. 11). During the intervention, the preparatory articles that the debaters read overlapped with the written cases in terms of their genre and communicative purpose (convincing someone of a particular standpoint). Because of the generic connection between the two modalities and because of the meaningfulness of reading in the debate environment (Mirra et al., 2016; Oros, 2007), linguistic and rhetorical forms could easily transfer from reading to writing. In the words of Ferris (2011), ‘reading gives students ideas and content to write about, models rhetorical strategies and genre specifications, and provides extensive input for acquisition of vocabulary and syntax occurring within authentic discourse’ (p. 161).

Debate pedagogy may also facilitate an effective interface between speaking and writing in a way that boosts writing development. In debates, the two output modalities can ‘mutually scaffold the transformation of complex, multidimensional thoughts into lines of spoken and written words’ (Belcher & Hirvela, 2008, p. 4). Previous research revealed that spoken interactions between learners provide scaffolding for their writing development (Hyland, 2008; Yang, 2008). In a collaborative dialogue through which students express ideas and reflect on what they and others say, learning takes place (Swain, 2000). The close and harmonious interplay of speaking and writing in the debate environment that facilitates a smooth transfer of gains from one modality to another has equally proved to be propitious for argumentative skills (i.e. communicative adequacy) (e.g. Chen et al., 2016).

As the components of debate task design are highly interconnected and coherent, gains seem to transfer, unhampered, from one modality to another. It stands to reason, therefore, to posit that many language forms that were encountered, for example, in

preparatory articles were employed in cases and then in actual debates. As these new gains featured in many debate cycles and even moved to a new cycle (i.e. new debate), their permanent retention stands a big chance. The debaters seemed to have benefited from these recursive cycles of debate, which enabled them to build sizable and readily accessible linguistic and argumentational resources that bolstered their argumentative essay writing.

Another inherent aspect of debate pedagogy that boosts language and argumentation development is its competitive environment. The urge to outshine opponents seems to put extra demands on debaters to produce adequate and sophisticated linguistic and rhetorical forms both during preparation (e.g. composing cases) and during actual debates. Debaters are aware of the fact that they are on a mission of defending their standpoints and outperforming their opponents. They seem to realize that this mission necessitates utilizing sophisticated, accurate, and adequate language. This endeavor prompts them to consciously employ linguistic and rhetorical forms that sophisticate different dimensions of argumentative writing.

In what follows, we delve further into the impact of the intervention on each writing area and explicate this effect in relation to debate pedagogy features.

1 Fluency

The participants in the intervention group comparatively produced longer essays after the intervention. This is a strong indication that these participants improved their argumentative writing abilities. Earlier research recognized text length as the most consistent and distinguishing indicator (Plakans et al., 2016) and ‘predictor of writing development and quality’ (Crossley, 2020, p. 416).

There are many relevant intervention-related factors that may have contributed to this significant improvement. One of these factors has to do with the positive impact of debating on argumentation competence. In a previous experimental study, we demonstrated that active participation in debates improved many aspects of argumentation, including argumentation quantity (El Majidi, et al., 2021b). Ease of generating arguments has seemingly had a marked impact on essay length. Another important factor concerns the reading-to-write pedagogy embedded in the intervention. The debaters seem to have benefited from the recurrent process of reading preparatory articles and subsequently composing related cases. These fruitful cyclic processes have apparently enabled the debaters to build up sizable vocabulary and to improve different aspects of language, including fluency (Hirvela, 2016; Hyland, 2019).

2 Syntactic complexity

The significant increase in two measures of complexity after the intervention indicates that the intervention was effective in helping learners produce more advanced and sophisticated language structures. These findings provide further evidence for the improvement of language quality after the intervention, as previous research associated longer T-units and more subordination (at the intermediate level) with higher writing

quality (e.g. Bulté & Housen, 2014). One factor that has probably induced these findings is related to the improvement of the structural complexity of argumentation (e.g. the use of elaborate arguments with warrants; El Majidi et al., 2021b). It appears that the debaters needed more complex syntactic structures to formulate complex arguments.

3 Lexical complexity

Though the measures of lexical complexity did not reach statistical significance, there is one interesting observation about lexical diversity whose mean increased in the intervention group. Previous research (e.g. El Majidi et al., 2018, 2020) showed that debaters improved their vocabulary after participating in L2 debates. The debate environment seems to promote the acquisition of new lexis and its active implementation. The debaters read preparatory articles that contained new words, many of which percolated into their cases and could then be transferred to their speaking discourse (in actual debates) and eventually reached their writing (see Malloy et al., 2020; Mirra et al., 2016). Each stage in this journey is expected to contribute to their permanent retention. Yet, more research is needed to assess the contribution of debate pedagogy to L2 lexical development more fully.

4 Accuracy

The findings also revealed that the intervention helped the debaters to produce essays with fewer errors. Although both groups improved their accuracy from pretest to post-test, the intervention group significantly outperformed the control group after the intervention. This is consistent with the findings of El Majidi et al. (2020) in which the debaters showed better improvement across many accuracy measures.

What might explain these results is the provision of systematic feedback in the debate intervention. Throughout the intervention, the intervention students received feedback on their written cases. Also, during debates the students were tasked to note down some of the mistakes their classmates made and to improve them. In addition, the instructors at times discussed some of the commonly made mistakes during actual debates. These repeated opportunities of drawing attention to form have presumably promoted the debaters' awareness of their gaps in linguistic knowledge, and they have accordingly refined them. The recursiveness of these feedback cycles and processing in each debate may have facilitated the transfer of gains from one debate to the next one and hence consolidated the gains.

Importantly, the debate environment attaches great importance to accurate output. It seems to infuse the debaters with the awareness that accurate language is needed to render arguments persuasive and ultimately outshine critical opponents. By contrast, inaccurate language is presumably perceived as impeding successful and effective persuasiveness. In a word, the debate environment prompts debaters to 'care about what they write' (Dickson, 2004, p. 35). Previous research showed that when L2 learners write with an authentic audience in mind, they tend to be more precise and accurate (e.g. Albadi, 2016; Cho & Choi, 2018).

5 Cohesion

The intervention group also made significant progress in producing more cohesive texts. After the intervention they wrote essays that displayed a greater number and a wider range of metadiscourse markers. This is an important finding, as organizational competence lies at the core of good-quality writing. Many studies found positive correlations between the use and diversity of metadiscourse markers and essay writing quality (Noble, 2010; Qin & Uccelli, 2016).

One possible explanation of these results might be ascribable to the fact that the debate context, by its nature, nurtures this area of performance. Composing good cases entails the use of a range of textual markers that establish effective connections between arguments. Therefore, the intervention students were asked to pay due attention to cohesive makers when constructing their written cases to mark smooth progression and shift from one argument to another. It stands to reason that cohesive markers transferred easily from debates to argumentative essay writing, as the two share the same genre, discourse, and mission. Furthermore, research showed that embedding cohesive markers in engaging, interesting, and challenging content promotes their advancement (Crosson & Lesaux, 2013).

6 Communicative adequacy

Beyond the linguistic dimensions, this research examined the effects of debate pedagogy on the communicative dimension of the written output, without which the assessment of writing proficiency in L2 is incomplete (Kuiken & Vedder, 2017, 2022). As the findings demonstrated, debate pedagogy exerted a sizable effect on all dimensions of this construct. After the intervention, the intervention group produced essays that contained more (counter)arguments (see the results of the dimensions of sufficiency of arguments and addressing the opposing view; see also the fluency measure). It seems that the debaters felt the need to produce a great deal of arguments to make their point of view more convincing. Also, the essays of the intervention groups were better organized (as suggested by the results of the organization of arguments dimension and also the cohesion measures). The debaters were also more able to produce well-structured essays with a flow of arguments that contribute to the achievement of the goal of the task (i.e. convincing someone of their point of view). In addition, as the results suggest, the produced arguments of the intervention group seem to be more comprehensible, elaborate, relevant, and more persuasive than the arguments presented in the essays of the control group. This indicates that the debate environment fostered the awareness that these argument qualities are needed to render their arguments (and eventually their essays as a whole) more convincing.

These findings are congruent with our hypothesis and in line with previous research that showed that debate pedagogy can effectively serve the argumentative discourse, which is at the core of the communicative adequacy of argumentative essays (e.g. Oros, 2007; Zorwick & Wade, 2016). For example, in the study of El Majidi et al. (2021b), the debaters comparatively increased their argumentation fluency, producing more (counter)arguments that were equally better organized and more persuasive than did their counterparts in the

control group. The debate environment seems to stimulate processes and provide incentives that orient the mind towards a sophisticated view of argumentation and develop metacognitive knowledge of argumentation.

Debate pedagogy involves dialogic and social negotiations that may enable debaters to obtain insights into the quality of their arguments (Chen et al., 2016; Matos, 2021). When provided with structured and authentic opportunities for meaningful dialogic interactions about appealing and engaging topics, students are prompted to craft clearer, well-structured, and well-reasoned written arguments (Kuhn & Crowell, 2011). These opportunities promote sophisticated reasoning that '[makes] its way into the writers' essays' (Kuhn et al., 2016, p. 105). Another merit of debate pedagogy that might have contributed to the current results is the presence of an audience (i.e. classmates, especially opponents, and the teacher). Berland and McNeill (2010) and Chen et al. (2016) contended that the audience gives students a strong impetus to develop rich, complex, and convincing arguments.

Debate pedagogy also seems to blend both learning-to-argue and arguing-to-learn perspectives, and argumentative essays profit from this blending (Zou et al., 2021). Through engagement in multiple debates, debaters may develop a good understanding of the architecture of argumentation and hence become better arguers (the learning-to-argue perspective). Equally, participation in debate facilitates exploring content from a variety of angles, enabling debaters to deepen their grasp of it (the arguing-to-learn perspective). The debate environment presumably enables the two argumentation perspectives to contribute synergistically to each other and eventually promotes argumentative writing. This means that improving the arguing skills may lead to the improvement of content mastery, and gaining a deeper understanding of content may empower arguing skills. Because of this, as an anonymous reviewer suggested, debate pedagogy may be useful not just for language classes but also for content classes, such as history.

VIII Limitations and conclusions

While promising, our results must be viewed in light of some limitations. The assessment of students' essay writing was based on two essays (pretest and posttest). Clearly a single essay at each time point may not be representative of the students' full essay writing ability. In addition, on the delayed posttest occasion, we only managed to elicit essays from two of the intervention classes. Therefore, caution needs to be exercised when interpreting the long-term effects of the intervention. Our study was further limited in terms of its sample, which only included students from *havo* and *vwo* tracks (relatively high-achieving students). The scope of future research should be broadened to include other tracks of secondary education (e.g. vocational students). Furthermore, this study focused on the effects of debate pedagogy on argumentative writing genre. It would be worthwhile investigating the extent to which this pedagogy can improve other writing genres. Last but not least, the assessment model (i.e. CAFAC measures) adopted in this study mainly focused on the linguistic production aspects. Communicative adequacy – which provides some insights into the content of the written production – investigated some quality aspects of argumentation that lead to task success. However, these quality aspects do not fully cover the argumentation construct, which is more complex and multilayered. In

order to obtain a comprehensive picture of debate-argumentation effects, we need to investigate other dimensions of argumentation, including the structural components of arguments (e.g. data, sub-arguments, warrants, backings, etc.).

Having acknowledged the limitations, we nonetheless believe that the findings of our study have yielded valuable new insights into the contribution of debate-based pedagogy to the development of L2 argumentative essay writing, an elusive valuable educational objective that has long been pursued. Our results suggest that debate pedagogy – like in our previous study involving opinion tasks (El Majidi et al., 2020) – creates an effective environment that coherently embeds activities that stimulate and scaffold processes that lead to the use of complex, accurate, coherent, and persuasive L2 language. The debate environment socializes the learning experience, promotes language awareness, demands the use of well-crafted and well-formulated arguments, and triggers linguistic and rhetorical processing (e.g. exposing gaps in L2 and reasoning skills) that promotes L2 learning. All these benefit argumentative essay writing in many ways.


It is noteworthy that the obtained effects in this study resulted from participating in 10 debates. It follows that instructors need to plan debates on a regular basis to obtain similar effects. Implementing debates in class may place extra demands on the instructor, but they are worth the time and effort.

The present findings have important implications for L2 argumentative essay writing pedagogy. We hope that they will lead to a reconceptualization of debate as a tool that mainly benefits oral and argumentation skills. As demonstrated in this study, the debate environment possesses many mechanisms and processes that stimulate the development of argumentative writing skills. We hope that the results of this study will encourage the introduction of debate into L2 language classrooms and equally provide impetus for investigating more pedagogical potential and affordances of debate pedagogy.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work has been financed by the Dutch Research Council (NWO; grant number 023.009.059).

ORCID iD

Abid el Majidi  <https://orcid.org/0000-0003-4840-2087>

Notes

1. In debates, 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). Students draw on cases during debates.
2. We thank an anonymous reviewer for bringing this point to our attention.
3. It is true that these delayed posttest results lack an important benchmark (i.e. a control group), yet they still provide relevant insights, albeit limited, into the durability of the obtained effects.
4. It is important to reiterate that the results of the delayed posttest should be interpreted with caution, since they only include two classes that both belong to the intervention group.

References

- Albadi, A.A.M. (2016). Twitter mobile application: A source of communicative and authentic learning. *Journal of Teaching English for Specific and Academic Purposes*, 4, 49–60.
- Belcher, D., & Hirvela, A. (Eds.). (2008). *The oral–literate connection: Perspectives on L2 speaking, writing, and other media interactions*. The University of Michigan Press.
- Berland, L.K., & McNeill, K.L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education*, 94, 765–793.
- Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, 26, 42–65.
- Chen, J. (2019). ‘To Blog, not to block’: Examining EFL learners’ language development and intercultural competence in the blogosphere through the sociocultural lens. In T. Dobinson & K. Dunworth (Eds.), *Literacy unbound: Multiliterate, multilingual, multimodal* (pp. 225–245). Springer.
- Chen, J.C., & Brown, K.L. (2012). The effects of authentic audience on English as a second language (ESL) writers: A task-based, computer-mediated approach. *Computer Assisted Language Learning*, 25(5), 435–454.
- Chen, Y.C., Hand, B., & Park, S. (2016). Examining elementary students’ development of oral and written argumentation practices through argument-based inquiry. *Science and Education*, 25, 277–320.
- Cho, Y., & Choi, I. (2018). Writing from sources: Does audience matter? *Assessing Writing*, 37, 25–38.
- Cohen, J. (1988). *Statistical power analyses for the behavioral sciences*. 2nd edition. Lawrence Erlbaum.
- Crossley, S. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, 11, 415–443.
- Crosson, A.C., & Lesaux, N.K. (2013). Connectives: Fitting another piece of the vocabulary instruction puzzle. *The Reading Teacher*, 67, 193–200.
- De Jong, N.H., Steinel, M.P., Florijn, A.F., Schoonen, R., & Hulstijn, J.H. (2012). The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performances of native and non-native speakers. In Housen, A., Kuiken, F., & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 121–142). John Benjamins.
- Dickson, R. (2004). Developing ‘real-world intelligence’: Teaching argumentative writing through debate. *The English Journal*, 94, 34–40.
- Dobbs, C.L. (2014). Signaling organization and stance: Academic language use in middle grade persuasive writing. *Reading and Writing*, 27, 1327–1352.
- El Majidi, A., de Graaff, R., & Janssen, D. (2018). Students’ perceived effect of in-class debates in second language learning. *The European Journal of Applied Linguistics and TEFL*, 7, 35–57.
- 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, 804–821.
- El Majidi, A., de Graaff, R., & Janssen, D. (2021a). Debate as a pedagogical tool for developing speaking skills in second language education. *Language Teaching Research*. Epub ahead of print 10 October 2021. DOI: 10.1177/13621688211050619
- El Majidi, A., Janssen, D., & de Graaff, R. (2021b). The effects of in-class debates on argumentation skills in second language education. *System*, 101, 102576.
- Ferretti, R.P., & Graham, S. (2019). Argumentative writing: Theory, assessment, and instruction. *Reading and Writing*, 32, 1345–1357.

- Ferris, D.R. (2011). *Treatment of error in second language student writing* (2nd ed.). University of Michigan Press.
- Ferris, D.R., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing, 10*, 161–184.
- Freeley, A.J., & Steinberg, D.L. (2005). *Argumentation and debate: Critical thinking for reasoned decision making*. 11th edition. Thomson & Wadsworth.
- Hirvela, A. (2013). Preparing English language learners for argumentative writing. In de Oliveira, L.C., & T.J. Silva (Eds.), *L2 writing in secondary classrooms: Student experiences, academic issues, and teacher education* (pp. 67–86). Routledge.
- Hirvela, A. (2016). *Connecting reading and writing in second language writing instruction*. 2nd edition. The University of Michigan Press.
- Hirvela, A. (2017). Argumentation and second language writing: Are we missing the boat? *Journal of Second Language Writing, 36*, 69–74.
- Huang, Y., & Jun Zhang, L. (2020). Does a process-genre approach help improve students' argumentative writing in English as a foreign language? Findings from an intervention study. *Reading and Writing Quarterly, 36*, 339–364.
- Hyland, F. (2008). Scaffolding during the writing process. In D. Belcher & A. Hirvela (Eds.), *The oral-literate connection: Perspectives on L2 speaking, writing, and other media interactions* (pp. 168–190). University of Michigan Press.
- Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. Continuum.
- Hyland, K. (2011). Learning to write: Issues in theory, research, and pedagogy. In Manchón, R.M. (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 17–35). John Benjamins.
- Hyland, K. (2019). *Second language writing*. 2nd edition. Cambridge University Press.
- Jin, T., Su, Y., & Lei, J. (2020). Exploring the blended learning design for argumentative writing. *Language Learning and Technology, 24*, 23–34.
- Jonassen, D.H., & Kim, B. (2010). Arguing to learn and learning to argue: Design justifications and guidelines. *Education Technology Research Development, 58*, 439–457.
- Kimura, M. (1998). Exploring relationships between oral and written communication: How speaking effects writing? *Eibeibunka: Studies in English Language, Literature and Culture, 28*, 23–35.
- Kuhn, D., & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents' thinking. *Psychological Science, 22*, 545–552.
- Kuhn, D., Hemberger, L., & Khait, V. (2016). *Argue with me: Argument as a path to developing students' thinking and writing*. 2nd edition. Routledge.
- Kuiken, F., & Vedder, I. (2017). Functional adequacy in L2 writing: Towards a new rating scale. *Language Testing, 34*, 321–336.
- Kuiken, F., & Vedder, I. (2022). Measurement of functional adequacy in different learning contexts: Rationale, key issues, and future perspectives. *TASK, 2*, 8–32.
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15*, 474–496.
- Lustigova, L. (2011). Speak your mind: Simplified debates as a learning tool at the university level. *Journal of Efficiency and Responsibility in Education and Science, 4*, 18–30.
- 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*, 79–100.
- Manchón, R.M. (2011). Situating the learning-to-write and writing-to-learn dimensions of L2 writing. In Manchón, R.M. (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 3–14). John Benjamins.

- Manchón, R.M., & Williams, J. (2016). L2 writing and SLA studies. In Manchón, R.M., & P.K. Matsuda (Eds.), *The handbook of second and foreign language writing* (pp. 567–586). De Gruyter Mouton.
- Matos, F. (2021). Collaborative writing as a bridge from peer discourse to individual argumentative writing. *Reading and Writing, 34*, 1321–1342.
- McKenney, S., & Reeves, T.C. (2012). *Conducting educational design research*. Routledge.
- McNamara, D.S., Crossley, S.A., & McCarthy, P.M. (2010). Linguistic features of writing quality. *Written Communication, 27*, 57–86.
- Mei, W.S. (2006). Creating a contrastive rhetorical stance: Investigating the strategy of problematization in students' argumentation. *RELC Journal, 37*, 329–353.
- Michel, M. (2017). Complexity, accuracy and fluency in L2 production. In Loewen, S., & S. Masatoshi (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 50–68). Routledge.
- Miller, J. (2008). *An introduction to English syntax*. 2nd edition. Edinburgh University Press.
- Mirra, N., Honoroff, B., Elgandy, S., & Pietrzak, G. (2016). Reading and writing with a public purpose: Fostering middle school students' academic and critical community literacies through debate. *Journal of Language and Literacy Education, 12*, 1–22.
- Noble, W. (2010). Understanding metadiscoursal use: Lessons from a 'local' corpus of learner academic writing. *Nordic Journal of English Studies, 9*, 145–169.
- Norris, J.M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*, 555–578.
- Oros, A.L. (2007). Let's debate: Active learning encourages student participation and critical thinking. *Journal of Political Science Education, 3*, 293–311.
- Ortega, L. (2011). Reflections on the learning-to-write and writing-to-learn dimensions of second language writing. In Manchón, R.M. (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 237–250). John Benjamins.
- Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics, 30*, 590–601.
- Pessoa, S., Mitchell, T.D., & Miller, R.T. (2017). Emergent arguments: A functional approach to analyzing student challenges with the argument genre. *Journal of Second Language Writing, 38*, 42–55.
- Plakans, L., Gebril, A., & Bilki, Z. (2016). Shaping a score: Complexity, accuracy, and fluency in integrated writing performances. *Language Testing, 36*, 161–179.
- Qin, W., & Uccelli, P. (2016). Same language, different functions: A cross-genre analysis of Chinese EFL learners' writing performance. *Journal of Second Language Writing, 33*, 3–17.
- Sasaki, M., Baba, K., Nitta, R., & Matsuda, P.K. (2020). Exploring the effects of web-based communication tasks on the development and transferability of audience awareness in L2 writers. *Australian Review of Applied Linguistics, 23*, 277–301.
- Snider, A., & Schnurer, M. (2006). *Many sides: Debate across the curriculum*. International Debate Education Association.
- Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *The Canadian Modern Language Review, 50*, 158–164.
- Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97–114). Oxford University Press.
- Swain, M. (2005). The output hypothesis: Theory and research. In Hinkel, E. (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–483). Lawrence Erlbaum.
- Takač, V.P., & Ivezic, S.V. (2019). Frame markers and coherence in L2 argumentative essays. *Discourse and Interaction, 12*, 46–71.

- Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In Ellis, R. (Ed.), *Planning and task performance in a second language* (pp. 239–273). John Benjamins.
- Turgut, Y. (2009). EFL learners' experience of online writing by PBWiki. In Siemens, G., & C. Fulford (Eds.), *Proceedings of ED-MEDIA 2009–World Conference on Educational Multimedia, Hypermedia, and Telecommunications* (pp. 3838–3847). Association for the Advancement of Computing in Education (AACE).
- Weissberg, R. (2006). *Connecting speaking and writing in second language writing instruction*. University of Michigan Press.
- Wingate, U. (2012). 'Argument!' helping students understand what essay writing is about. *Journal of English for Academic Purposes*, 11, 145–154.
- Wright, K.L., Hodges, T.S., Enright, E., & Abbott, J. (2021). The relationship between middle and high school students' motivation to write, value of writing, writer self-beliefs, and writing outcomes. *Journal of Writing Research*, 12, 601–623.
- Yang, L. (2008). From group talk to group writing. In D. Belcher & A. Hirvela (Eds.), *The oral-literate connection: Perspectives on L2 speaking, writing and other media interactions* (pp. 139–167). The University of Michigan Press.
- Yasuda, S. (2019). Children's meaning-making choices in EFL writing: The use of cohesive devices and interpersonal resources. *System*, 85, 1–15.
- Yoon, H.J., & Polio, C. (2017). The linguistic development of students of English as a second language in two written genres. *TESOL Quarterly*, 51, 275–301.
- 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, 434–444.
- Zou, M., Li, X., & Lee, I. (2021). Blending learning to argue and arguing to learn in EFL writing instruction: A classroom inquiry. In Hirvela, A., & D. Belcher (Eds), *Argumentative writing in a second language: Perspectives on research and pedagogy* (pp. 169–186). University of Michigan Press.