Debate as L2 Pedagogy: The Effects of Debating on Writing Development in Secondary Education

ABID EL MAJIDI,[1] RICK DE GRAAFF,[2] AND DANIEL JANSSEN[3]
[1] *Utrecht University, Department of Languages, Literature and Communication – Linguistics, Trans 10, 3512 JK, Utrecht, the Netherlands. Email: a.elmajidi@uu.nl*
[2] *Utrecht University, Department of Languages, Literature and Communication – Linguistics, Trans 10, 3512 JK, Utrecht, the Netherlands. Email: r.degraaff@uu.nl*
[3] *Utrecht University, Department of Languages, Literature and Communication – Linguistics, Trans 10, 3512 JK, Utrecht, the Netherlands. Email: D.M.L.Janssen@uu.nl*

Research has painted a pessimistic picture of students' second language (L2) writing skills in secondary education. One innovative tool that may help students foster their L2 proficiency, including writing ability, is in-class debate. Debate is commonly associated with oral communication and argumentation skills. However, debate may also offer advantages as an effective vehicle for L2 writing instruction. This study evaluates the effect of a debate intervention on the writing competence 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 design with a control group, involved 8 classes at 3 secondary schools in the Netherlands ($N = 146$). To measure the effect of the intervention, we analyzed 2 opinion writing tasks produced by the students: just prior to the first debate (pretest) and approximately 10 weeks later (posttest). We used a variety of measures, tapping into different aspects of writing performance, including fluency, syntactic and lexical complexity, accuracy, and cohesion. Multilevel analysis of the data revealed that the students in the intervention group significantly improved on a number of measures in comparison to the control group.

Keywords: in-class debate; writing development; second language acquisition; secondary education; L2 writing instruction

WRITING EFFECTIVELY IS REGARDED AS A vital 21st-century skill for personal, academic, and professional success (Allen, 2018). With the rapid growth of globalization and the importance of international correspondence, the need to be proficient in the English language in general, and in English writing in particular, has increased (Naghdipour, 2016). Almost all forms of personal, academic, and business communication are now carried out in English and, evidently, the more proficiently we communicate, the more successful we can be in life and at work. In the Netherlands, this awareness has led to interesting developments in the educational system: a growing number of multilingual secondary schools (English and Dutch) and a growing number of universities that offer bachelor's and master's programs in English only.

Naghdipour (2016) and Allen (2018) pointed out that second language (L2) writing instruction has not effectively responded to the increasing significance of writing in English. A number of studies have painted a pessimistic picture of the (L2) writing skills of secondary-school students in the Netherlands (Beeker et al., 2015) and elsewhere (Leki, Cumming, & Silva, 2010), and

The Modern Language Journal, 104, 4, (2020)
DOI: 10.1111/modl.12673
0026-7902/20/804–821 $1.50/0

the effects thereof on society and higher education. For example, Beeker et al. (2015) reported that only 51% of Dutch secondary advanced students manage to achieve the target level (B2 in the Common European Framework of Reference for Languages [CEFR], Council of Europe, 2018). Naghdipour (2016) and Polio and Park (2016) called for more interventions to inform L2 writing pedagogy about how to help students write effectively.

Writing is a highly demanding task that requires cognitive processing, enactment of linguistic knowledge, awareness of the social context underlying the written communication (Qin & Uccelli, 2016), and a higher level of precision than speech (Manchón & Williams, 2016). Seeing its complexity and the current worrying situation, "we need to search for innovative pedagogical tools and strategies to respond more effectively" to these challenges (Uccelli, Dobbs, & Scott, 2013, p. 37). One of these innovative tools that may help students foster their L2 writing ability is debate. Debate is widely acclaimed as an effective teaching tool and is believed to hold promise as a conducive mechanism for L2 learning (e.g., Lustigova, 2011; Zare & Othman, 2013). Debate is commonly associated with oral communication and the skill of argumentation. It is our view that debate also offers advantages as an effective vehicle for L2 writing instruction. It affords learners an opportunity to practice and attend to language processes that help improve L2 writing.

We know of little research into the benefits of debating for improving L2 writing skills, especially in secondary schools. The main aim of this study is therefore to evaluate the effect of debating on the writing competence of Dutch secondary-school students. To this end, we conducted a debate intervention comprising writing components in addition to actual debates, including summarizing preparatory articles, making notes, and writing cases.[1] We assessed the debate-writing effect by measuring effects on the linguistic features of opinion-writing tasks in a pretest–posttest design.

SECOND LANGUAGE ACQUISITION AND L2 WRITING INSTRUCTION

Manchón and Williams (2016) remarked that, in recent years, increasing effort has been invested in cross-disciplinarity and the development of second language acquisition (SLA)–L2 writing interfaces. They identified three areas where L2 writing research and SLA can overlap: "1. the development of learners' written language over time; 2. the contribution of general L2 proficiency to writing; and 3. the contribution of writing, writing instruction, and feedback to L2 proficiency" (p. 568). The current research promotes this interface by making some modest contributions, in particular to the third area.

Writing instruction can, broadly speaking, be product-oriented or process-oriented. While the product-oriented approach sees the written product as an end in itself, the process-oriented approach puts emphasis on the subject matter, ideas, and the negotiation of meaning (Badger & White, 2000) and highlights the writing process and procedures that learners need to go through before producing a written product, including drafting, revising, and editing. Yong (2010) contended that process writing offers learners an enriched learning experience, enhances their interest in the subject matter, and stimulates their critical thinking. Writing in our debate intervention seems to subscribe to process-oriented instruction: For example, writing a case foregrounds the negotiation of arguments and involves recursive processes of composing, revision, and editing.

Manchón (2011) identified three L2 writing perspectives that are relevant to L2 writing instruction: learning to write (LW), writing to learn content (WLC), and writing to learn language (WLL). In the LW perspective, writing is learned and taught as an end in itself. The WLC perspective sees L2 writing as a medium for learning disciplinary subject matter in the content areas. The WLL perspective, which constitutes an important area of study in the research agenda on SLA–L2 writing interfaces (Manchón & Williams, 2016), regards writing as a means of promoting language learning mainly through offering learners the opportunity to improve their own writing by guiding them through the revision process to raise their awareness of problematic linguistic areas in their output.

Feedback plays a central role in the three writing perspectives. It is generally regarded as an important instrument for scaffolding L2 writing in that it holds the potential to raise consciousness and noticing in the learning process by directing learners' attention to the problematic forms in their output (Hyland, 2011). Becoming aware of these problematic areas is considered an essential condition for language learning to take place (Schmidt, 2001).

As we shall see, the three writing perspectives in our debate intervention are blended and reinforce one another (especially the LW and WLL perspectives). Ortega (2011) contends that the interface between these writing perspectives (LW, WLC, and WLL) can lead to synergistic benefits for L2 writing development. In debates,

students engage in writing (i.e., case writing) with a communicative purpose, that is, defending their proposition and weakening that of their opponent. This orientation stimulates students to focus on negotiating meaning. Here, writing operates as a vehicle for synthesizing and analyzing arguments (the WLC perspective). This approach also corresponds with Hyland's (2011) view of a successful LW implementation, wherein a text is seen as a social and reader-oriented discourse that conveys the writer's intentions, ideas, and perspectives.

Crafting a persuasive case also entails the use of accurate and sophisticated language. In our debate intervention, reading-to-write pedagogy (e.g., Hirvela, 2016) and the processing of feedback promote this focus on language. Before embarking on writing each case (see the Intervention section), students read two articles. This allows them, for example, to learn new words and immediately employ some of them in the subsequent case. In other words, "reading … provides writers with essential material to write with and about" (Hirvela, 2016, p. 49). Importantly, reading and producing texts that share the same genre facilitates the transfer of linguistic forms from read texts to written texts (Hyland, 2007; So, 2005). Earlier research has revealed that genre pedagogy that integrates reading and writing is fruitful for L2 (Yang, 2016) and foreign language (FL) writing (Shum, Tai, & Shi, 2018). Furthermore, debating enables a cyclic processing of feedback: For each case, the student goes through the process of writing, feedback processing, and revision. This cycle enables students to go through linguistic processing (e.g., noticing gaps in their L2) that promotes learning, hence refining their knowledge of the L2 (Manchón, 2011). In such a context, learners make form–meaning connections; as a result, writing operates as a facilitating factor, functioning as a vehicle for learning (the WLL perspective; Williams, 2012).

THEORETICAL CONSIDERATIONS BEHIND DEBATE AS L2 WRITING PEDAGOGY

We will discuss two relevant theoretical perspectives—namely, Swain's (1993) Output Hypothesis and Long's (1996) Interaction Hypothesis—that corroborate our idea that debating may be an effective pedagogical framework for L2 writing development.

Swain (1993) argued that L2 learners "push their linguistic competence to its limit as they attempt to express their ideas" (p. 162). In her view, output pushes learners to process language more deeply and effectively than processing language through reading and/or listening alone. Research has suggested that it is mainly these output processes that lead to L2 development (Manchón & Williams, 2016). The act of writing, by its nature, involves a greater need and opportunity to focus on form than does speaking; the slower pace of writing offers learners more freedom and space to reflect critically on both content and form (Manchón & Williams, 2016).

Stewart (2003) suggested that debates can generate a great deal of output as debaters hold different views and need to sell their standpoints. The inherently competitive environment of debate—what is more compelling than outsmarting your fellow classmates?—provides students with a functional context in which they are able to produce functional output in both oral contributions and written ones. Furthermore, debate enables learners to engage in bidirectional output exchange, which allows them to compare their output with each other, notice gaps in their L2, and hence attend to and remedy the problematic areas in their interlanguage (Swain, 1993). Under such circumstances, output production will stimulate L2 development in general and writing in particular.

A second theoretical perspective that supports the idea that debating may be an effective tool for language learning is the Interaction Hypothesis (Long, 1996). This theory suggests that learners develop their L2 when they engage in negotiating meaning. Several studies have shown that student interaction in the learning process is an impactful variable that facilitates language acquisition. For example, Pica, Kang, and Sauro (2006) argued that student interaction activates attentional processes and facilitates attention to form, function, and meaning. The authors also suggested that through attentional processes, L2 learners become aware of the shortcomings in their input, thereby modifying their output.

In-class debates involve rich and multilevel interactions that make it easier for language learners to notice gaps and hence reflect on and revise their L2 output. First, students interact with content as they read, select, and arrange information, arguments, and texts. Second, students interact with the instructor as they provide feedback on their performance. Finally, students interact with each other. Seen in this perspective, it is clear why Wade (1998) lauded the interactive pedagogical merit of debate when he stated that "there are certainly trends in education which encourage interactive and dialogic pedagogies, but few are as potent as debate" (p. 63). It is true that most interactions and meaning negotiations in debating take place during actual debates, but research (e.g., Cho, 2017) has shown that writing

also benefits from these oral interactions and negotiations.

The close connection between speaking and writing facilitated by debate can reinvigorate writing development. The debate environment promotes a smooth and recursive movement between the two skills in a way that stimulates transfer of gains to move from one mode to the other. Previous research has revealed that spoken interactions between learners provide scaffolding for their writing development (e.g., Yang, 2008). In such an environment, we can assume that the two modes can "mutually scaffold the transformation of complex, multidimensional thoughts into lines of spoken and written words" (Belcher & Hirvela, 2008, p. 4). This mutual and close conjunction of speaking and writing has also proved to be propitious to the produced content—namely, argumentative skills (e.g., Chen, Hand, & Park, 2016; Chen, Park, & Hand, 2016).

In the debate context, composition is viewed as a socially oriented activity: Participants produce cases for a particular purpose and aimed at a specific audience (teacher and classmates, especially opponents). Chen, Hand, et al. (2016) and Chen, Park, et al. (2016) stated that the audience provides students with additional motivation to develop rich and convincing reasoning. They further suggested that the audience is a critical factor that inspires students to develop persuasive and more complex arguments, and that helps them to connect oral and written arguments. In the same vein, Turgut (2009) found that the presence of a real audience stimulated the development of writing skills among FL learners, who paid more attention to dimensions of language use such as word choice.

Another factor that may offer students an extra impulse to take the learning process more seriously during debates is their positive attitude toward debating as a pedagogical tool (e.g., el Majidi, de Graaff, & Janssen, 2015, 2018; Lustigova, 2011). For example, Lustigova (2011) reported that 75% of the students participating in a debate course in the first semester continued into the second semester and thus participated in debate sessions for an entire academic year. Recent empirical research has revealed that there is a positive correlation between task attitude and language acquisition (Dewaele et al., 2018).

THE PRESENT STUDY

Very few studies in either the L1 or the L2 context have examined the effect of debate on language performance in general and on writing performance in particular. The studies that have explored this effect in L2 or FL so far were based on observations (Lustigova, 2011), interviews (Aclan & Aziz, 2015), or questionnaires (mainly focused on speaking; Zare & Othman, 2015); most were aimed at students in higher education. One of the studies involving L2 secondary-school students was conducted by el Majidi et al. (2018). The participants in this study perceptively correlated debate participation with an improvement in writing with a mean of 3.94 on a 5-point Likert scale.

To the best of our knowledge, no empirical studies have been conducted in L2 or FL research to demonstrate to what extent in-class debates improve students' writing proficiency across different areas of performance. Polio and Park (2016) called for the effect of interventions on different linguistic aspects of writing to be charted, including fluency, syntactic and lexical complexity, and accuracy. Following this call, the present study explores the impact of in-class debates on L2 writing development among secondary school students—an understudied population in L2 writing research, according to Qin and Uccelli (2016)—with respect to fluency, syntactic and lexical complexity, accuracy, and cohesion. More specifically, this study will address the following research question:

RQ. What are the effects of in-class debate activities on different aspects of L2 writing proficiency, including fluency, syntactic complexity, lexical complexity, accuracy, and cohesion among Dutch secondary-school students?

H. Given that learners in the debate intervention were involved in a rich writing environment that stimulates processes that promote L2 writing development, they were likely to compose texts of better quality across various measures than control-group students.

To answer this question, we conducted an intervention with a pretest–posttest control group quasi-experimental design. The main source of data was opinion-writing tasks (in which the students argued for or against a controversial topic) produced by the students on two occasions: just prior to the first debate (pretest) and approximately 10 weeks later, prior to the tenth debate (posttest).

METHOD

Participants

The study sample consisted of eight intact classes at three secondary schools in the

Netherlands ($N = 146$) where English is taught as a L2. These schools are located in three urban areas (Rotterdam, Leiden, and Alkmaar) with comparable graduation rates. Five classes were in their fifth year of the higher general secondary education track (*havo 5*)[2] ($n = 88$) and three classes were in their fourth year of the secondary preuniversity education track (*vwo 4*) ($n = 58$). Five classes served as the intervention group ($n = 95$), and three as the control group ($n = 51$). Six classes (three intervention classes and three control classes) came from one school, with the first author as the instructor, and the other two intervention classes came from the other two schools (two different instructors). The participants included 87 females and 59 males, ranging in age from 15 to 18. The English proficiency level (including writing) of all classes spanned mostly the B1 (the third level of English in CEFR; this level is comparable to the intermediate level) and B2 levels (the fourth level of English in CEFR; this level is comparable to the upper intermediate level), as estimated by their teachers. With the exception of one intervention class 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 (reading, writing, listening, and speaking). For the purposes of this study, while the intervention students were involved in the debate intervention (once a week), the control students received extra regular instruction (during that same session) during which the four language skills were further practiced. With regard to their writing experience in previous years, the participants received the same writing instruction, which was not intensive. Writing was mostly covered in tasks from coursebooks, which mostly require students to produce only short texts.

Design

As mentioned, we opted for a pretest–posttest control group quasi-experimental design. The intervention group participated in 10 debates (one debate per week), which were part of the class curriculum, each lasting approximately 50 minutes. To enhance the external validity of the findings, we gathered data from three different secondary schools. The participating teachers were trained and familiarized with the content of the intervention by the first author.

Intervention

Our debate task design was validated in a previous study following the principles and guidelines of educational design research. Prior to each debate, the students received preparation time of 1 week. They were allowed to choose the topics they found interesting (e.g., euthanasia; the right to bear arms), and they were free to choose their side as protagonist or antagonist. We dedicated one session to each topic, and we used two debate formats: 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).

Each debate consisted of three stages: predebate, debate, and postdebate. Table 1 presents the writing activities performed in each debate session, in addition to the writing tasks performed by the control group during the intervention.

It is important to note that the intervention students also wrote persuasive essays and letters, but fewer than the control students. More precisely, during the intervention, the control group wrote on average two more essays and one more letter than the intervention group. In addition, both groups received brief instruction on how to compose essays and letters and received feedback on their produced texts. Neither group was trained in the specific task we used in the intervention—an opinion task—but both groups had some experience with composing texts in which they had to formulate a standpoint and support it with arguments.

Procedures

To measure the effects of our debate intervention on writing proficiency, we compared two free opinion tasks (as pre- and posttests). We selected two (controversial) topics: (a) capital punishment should be legalized (henceforth, capital punishment); and (b) abortion should be banned (henceforth, abortion). Both topics are accessible and of interest to students in this age group. The topics of the pre- and posttest opinion tasks (capital punishment and abortion) were not discussed beforehand and were identical for all participating classes, but we counterbalanced the order of their administration to avoid any potential topic effect (including control group). The conditions around pre- and postassessments were the same for both the intervention and control groups. In both groups, students did not receive a grade for these tasks.

TABLE 1
Relevant Activities Performed During the Intervention

Intervention students	Control students
Predebate stage	Reading and summarizing newspaper articles (e.g., *The Guardian*) covering current issues, including argumentative articles addressing, for example, issues related to politics and changes in policies
Reading two articles and summarizing them. The instructor provided one article and the students had to find another	
Writing a case	
During-debate stage	Writing persuasive essays on controversial subjects (e.g., school uniforms should be introduced in schools)
Noting down the arguments of the opponents, as debaters have to rebut them during the rebuttal and clash stages	
Noting down new words and mistakes from classmates	
Postdebate stage	Writing letters, especially complaint letters in which students have to express dissatisfaction with a particular service and accordingly convince the addressed company to provide a refund
Processing the feedback provided by the instructor on written cases	

We opted for a free-opinion-writing task because it lowers the threshold for expressing one's ideas or point of view. Unlike many other forms of argumentative writing, free opinion writing is not constrained by genre conventions and hence its composition is likely to be more accessible to language learners. Dobbs (2014) contended that opinion tasks[3] are accessible to middle schoolers and yield a representative picture of a learner's writing proficiency. Likewise, Hirvela (2017) held that effective argumentation (which lies at the heart of opinion tasks) is an important indicator of L2 writing ability. Furthermore, opinion tasks lend themselves more readily to the elicitation and hence assessment of cohesion than, for example, narrative tasks. It is important to mention that both the intervention and the control groups practiced different forms of argumentative writing (cases, essays, requests for refunds) that can be seen as adequate and more or less equivalent preparation for the opinion task used to elicit data for the study.

Before performing the opinion tasks in class, the students in both groups received 25 minutes of preparation time (Qin & Karabacak, 2010). They received a preselected article with opposing views and were allowed to search the Internet for more arguments. After 25 minutes, we collected the articles to prevent the students from plagiarizing. Then, the students had 15 minutes to write down as many arguments as possible in support of their standpoint. No further instruction was given. We obtained consent from parents to use their children's data for research purposes.

Measures

To assess the quality of the participants' written tasks in the pretest and posttest, we used a variety of genre-independent measures, tapping into different aspects of performance. Given the multidimensional nature of L2 proficiency and development, it was essential to select a range of measures to track relevant differences in performance between pretest and posttest (Wolfe–Quintero, Inagaki, & Kim, 1998). For this reason, we analyzed the texts produced for indicators of fluency, syntactic and lexical complexity, accuracy, and cohesion. The measures were a mixture of automatically coded features and measures that required hand coding.

Fluency. Fluency was measured in terms of the total number of words produced in 15 minutes. This is the most common metric for measuring written linguistic production (Plakans, Gebril, & Bilki, 2016), and has been shown to differentiate between different levels of writing ability and to capture L2 learners' writing development (Sasaki, 2004).

Syntactic Complexity. As measures of syntactic complexity, we used three indices recommended by Norris and Ortega (2009) that capture the multifaceted nature of this construct:

1. global complexity (number of words per T-unit, MLT)
2. complexity by subordination (mean number of clauses per T-unit, C/T)

3. clausal/phrasal complexity (mean length of clauses, MLC)

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

Accuracy. To measure accuracy, we first segmented the written texts (292 in total) into clauses. Following Miller's (2008) overview, nonfinite clauses were also coded as subordinate clauses because "they express propositions and, like finite clauses, consist of a verb plus complements and adjuncts" (p. 85). Unfinished sentences at the end of texts were not coded for errors. We calculated the following indices:

1. error-free clauses (EFCs)
2. lexical errors per 100 words (lex.er/100)
3. syntactic errors per 100 words (syn.er/100)
4. morphological errors per 100 words (mor.er/100)
5. prepositional errors per 100 words (pre.er/100).

We used the ratio of EFCs as a general measure, since it is widely recognized as a reliable measure for tracking changes in accuracy (e.g., Tavakoli & Skehan, 2005). In addition, we calculated the number of errors per 100 words (Inoue, 2016; Mehnert, 1998). However, since different linguistic categories (lexical, syntactic, morphological, and prepositional) represent separate knowledge domains (Ferris & Roberts, 2001), we computed separate ratios for these grammatical categories (see Yoon & Polio, 2017, for the operationalization and examples of the first three measures).[4] Furthermore, we ignored spelling, spacing, and punctuation errors, unless a misspelled word resulted in an actual English word (Ferris & Roberts, 2001).

Lexical Complexity. To track the effect of the debate intervention on the development of lexical proficiency, we used two measures of lexical sophistication and one measure of lexical diversity obtained from the computational tool Coh-Metrix (McNamara, Crossley, & McCarthy, 2010):

1. average word length (WL)
2. word frequency (WF)
3. measure of textual lexical diversity (MTLD)

WL has been widely employed as an approximation of lexical sophistication and is regarded as an effective predictor of sophisticated vocabulary, with longer words indicating more sophistication (Yoon, 2017; Yoon & Polio, 2017). The WF index calculates the mean logarithmic frequency for all words. It describes how often particular words occur in the English language, drawing on the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995). A lower WF thus indicates higher sophistication. We used MLTD to measure lexical diversity. In addition to the fact that MTLD is less affected by text length, it "also allows for comparisons between text segments of considerably different lengths (at least 100 to 2000 words) and produces reliable results over a wide range of genres" (McNamara et al., 2010. 69).

Cohesion. To convey an organized sequence of ideas and arguments, students need to employ cohesive devices. In this study, we measured cohesion by tracking the incidence of connectives, which play a key role in creating cohesive links between ideas and clauses (Halliday & Hasan, 1976).

We adopted Hyland's (2005) framework for interactive metadiscourse as an analytic framework to assess interactive metadiscourse markers, that is, words and phrases used to explicitly signal the coherent organization of ideas and arguments in a text to guide readers. Coh-Metrix also enables an automated computerized analysis of cohesion indices. However, earlier research (Uccelli et al., 2013) has alluded to the point that Coh-Metrix might not accurately measure cohesion, as it might fail to distinguish between, for example, cases in which certain organizational markers function as cohesive markers and other cases in which they fulfill other functions, as is the case in this sentence: *I like this too so much.* In this sentence, Coh-Metrix would count *too* and *so* as cohesive markers, while they are not. Hyland's analytic framework seems to provide a more fine-grained analysis of cohesive devices, since they are manually identified and each marker is evaluated in its own right. Furthermore, research has shown that the frequency and diversity of these metadiscourse markers significantly reflect the quality of argumentative or persuasive texts (Qin & Uccelli, 2016; Uccelli et al., 2013).

Following Hyland's procedures, we coded the produced texts for four types of organizational marker, in addition to their diversity of type and token (see, for example, Dobbs, 2014; Qin & Uccelli, 2016; Uccelli et al., 2013):

1. frame markers: markers that mark the sequence of arguments or counterarguments (e.g., *firstly*, *secondly*)

2. code glosses markers: markers that introduce an example or paraphrase (e.g., *for instance*, *in other words*)
3. 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)
4. conclusion markers: markers that introduce a summary or conclusion (e.g., *in conclusion*, *all in all*)
5. markers diversity token: diversity of markers in terms of token
6. markers diversity type: diversity of markers in terms of type

Interrater Reliability

To assess interrater reliability for the hand-coded measures, a randomly selected sample of 25% of the total data was verified by a research assistant, who was blind to the intervention condition of participants. Cohen's Kappa was high for the assessed measures: .98 for clause identification, .86 for mistake identification, .88 for mistake categorization, .92 for EFC identification, and .95 for discourse markers. Disagreements were resolved through discussion until complete agreement was achieved.

Statistical Analysis

As our participants came from different classes within different schools, our data were structured hierarchically. The performance of students within one class or school is likely to be more homogeneous and therefore more analogous to each other than to the performance of students in other classes and schools. We therefore applied multilevel linear analysis (MLA). The multilevel procedure enabled us to explicitly model possible dependencies in the data. In this study, we used a two-level hierarchical linear model to account for the multilevel data structure, with students nested within classes. We modelled 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. posttest) with group (intervention vs. control group). For these reasons, we limit ourselves to reporting interactions. Prior to performing statistical analyses, we checked the prerequisite assumption that the different residual scores are normally distributed by a visual analysis of the histograms of each residual, which is the standard procedure in multilevel modelling. No notable deviations were visible.

RESULTS

The objective of this study was to evaluate the effectiveness of in-class debates on the written performance of secondary school students. To this end, we obtained two scores (pretest and posttest) for each learner for texts produced in each condition (intervention and control groups) and for each measure (see the Appendix for an example of the analysis of a text produced by a participant). Table 2 presents the descriptive statistics (estimated means and standard errors) for the performance of the students in both the control and intervention groups. The figures indicate that some measures generally improved over the intervention period. It seems that the intervention group performed better than the control group across the majority of outcome variables at posttest.

To estimate the magnitude of the difference between the intervention and control groups, we used Cohen's d effect size when significant differences were observed. The MLA results shown in Table 3 are presented in relation to the measures discussed in the previous section.

As for fluency, the results indicate that the participants in the intervention group were able to produce significantly longer texts than their counterparts in the control group, $F(1,284.136) = 4.5$, $p = .017$. On average, the debaters produced 40 more words in the posttest compared to the pretest, while the control participants produced 6 words more at posttest. The effect size d of this improvement is moderate, $d = .44$.

Concerning syntactic complexity, the students in the intervention group showed a significant improvement in terms of clause length, $F(1,292) = 5.4$, $p = .011$, with a moderate effect size, $d = .57$. The other two measures, however, fell short of significance.

Regarding lexical complexity, it was found that the mean word length, $F(1,284.359) = 4.1$, $p = .022$, was significantly different; again, the intervention group outperformed the control group, and the effect size of the difference was moderate, $d = .52$. As for word frequency and lexical diversity, the intervention group also improved more than the control group from pretest to

TABLE 2
Means (and Standard Errors) of Outcome Variables Across Time and Condition

		Intervention ($n = 95$)				Control ($n = 51$)			
Measure	Index	Pretest		Posttest		Pretest		Posttest	
Fluency	Number of words	167.89	(18.47)	208.30	(18.47)	190.44	(24.01)	197.17	(24.01)
Syntactic complexity	MLT	13.33	(.58)	13.58	(.58)	14.22	(.76)	13.76	(.76)
	MLC	7.27	(.12)	7.76	(.12)	7.59	(.16)	7.43	(.16)
	C/T	1.85	(.07)	1.76	(.07)	1.88	(.09)	1.85	(.09)
Lexical complexity	MLTD	64.41	(1.69)	68.92	(1.69)	59.65	(2.31)	60.79	(2.31)
	Word frequency	3.16	(.01)	3.13	(.01)	3.15	(.02)	3.15	(.02)
	Word length	4.12	(.03)	4.27	(.03)	4.18	(.04)	4.20	(.04)
Accuracy	EFCs	0.66	(.02)	0.75	(.02)	0.69	(.02)	0.72	(.02)
	Lexical errors	1.65	(.21)	0.87	(.21)	1.36	(.28)	1.21	(.28)
	Syntactic errors	0.99	(.11)	0.54	(.11)	0.75	(.15)	0.57	(.15)
	Morphological errors	4.26	(.33)	3.10	(.33)	3.50	(.44)	3.42	(.44)
	Preposition errors	0.41	(.07)	0.29	(.07)	0.53	(.09)	0.41	(.09)
Cohesion	Transition markers	4.89	(.67)	6.23	(.67)	5.90	(.87)	6.49	(.87)
	Frame markers	0.71	(.19)	1.91	(.19)	1.10	(.25)	0.65	(.25)
	Gloss markers	0.52	(.14)	0.66	(.14)	0.53	(.18)	0.55	(.18)
	Conclusion markers	0.10	(.04)	0.45	(.04)	0.12	(.06)	0.20	(.06)
	Diversity type	1.80	(.11)	2.67	(.11)	1.98	(.14)	1.88	(.14)
	Diversity token	4.23	(.52)	7.22	(.52)	5.36	(.68)	5.03	(.68)

Note. 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; EFCs = error-free clauses.

TABLE 3
Multilevel Analysis Results

		Fixed effects				Random effects	
Measures	Index	Denominator df	*F*	*P* (one-tailed)	*d*	Variance within class	Variance between classes
Fluency	Number of words	284.136	4.523	.017	0.44	4162.40	1477.48
Syntactic complexity	MLT	283.872	.949	.166		8.96	1.18
	MLC	292	5.409	.011	0.57	1.31	.00
	C/T	284.043	.560	.228		.13	.01
Lexical complexity	MLTD	292	.690	.203		272.77	.00
	Word frequency	283.783	1.542	.108		.01	.00
	Word length	284.359	4.143	.022	0.52	.06	.00
Accuracy	Error free clauses	283.294	3.112	.040	0.41	.02	.00
	Lexical errors	282.527	4.078	.022	0.48	1.60	.14
	Syntactic errors	282.983	1.978	.081		.64	.03
	Morphological errors	284.575	3.474	.032	0.45	5.54	.24
	Preposition errors	285.315	.000	.498		.31	.00
Cohesion	Transition markers	284.226	.939	.167		9.90	1.68
	Frame markers	284.151	32.557	.000	1.35	1.39	.10
	Gloss markers	284.244	.397	.133		.57	.06
	Conclusion markers	292	7.137	.004	0.63	.18	.00
	Diversity type	284.394	28.190	.000	1.27	.56	.28
	Diversity token	284.018	37.492	.000	1.36	4.89	1.07

Note. 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.

posttest (see Table 2); however, this improvement fell short of significance.

With regard to accuracy, we found that the debate intervention affected various aspects of accuracy to different degrees. Although learners in both groups displayed improvement across almost all accuracy measures, MLA revealed that the greater improvement was experienced by the intervention group. It was found that the intervention group significantly improved in terms of EFCs, $F(1,283.294) = 3.1$, $p = .040$, with a moderate effect size, $d = .41$. The analysis also revealed significant improvement in favor of the intervention group in terms of lexical, $F(1,282.527) = 4.1$, $p = .022$ and morphological errors, $F(1,284.575) = 3.5$, $p = .032$, with moderate effects on both indices, $d = .48$ for lexical errors and .45 for morphological errors. No significant differences were found in syntactic errors and preposition errors. Note that the majority of the errors made in both groups concerned morphological and lexical errors, and that this distribution of errors remained stable over time.

As for cohesion, MLA revealed that the intervention group significantly improved compared to the control group in terms of the use of frame markers, $F(1,284.151) = 32.6$, $p < .001$, with a very large effect size, $d = 1.35$; conclusion markers, $F(1,292) = 7.1$, $p = .004$, with a moderate to large effect size, $d = 0.63$; diversity of type, $F(1,284.394) = 28.2$, $p < .001$, with a very large effect size, $d = 1.27$; and diversity of token, $F(1,284.018) = 37.5$, $p < .001$, also with a very large effect size, $d = 1.36$.

An important note should be made regarding the use of transition markers. Though both groups improved from pretest to posttest, with the intervention group improving more, the intervention group used far more sophisticated transition markers at posttest, as Table 4 displays.

DISCUSSION

The main goal of the present study was to assess the impact of an in-class debating intervention on L2 learners' written performance. Results showed that L2 learners of the intervention group significantly improved their writing performance, displaying an increase in fluency and a number of measures of syntactic and lexical complexity, accuracy, and cohesion after the intervention. These findings indicate that the use of in-class debates as a teaching tool is indeed effective in improving L2 writing in secondary education.

The debate intervention offered students opportunities to practice writing before and during debates. The case was written and rewritten in an attempt to convey arguments in concise, clear, and powerful language that would convince opponents during debates. In addition, we believe that the repetitive process of case writing (prior to each debate) created the opportunity for students to carry over what they gained from one debate to the next (in terms of language development), and hence reinforce these gains. This finding is in line with previous research that showed that learners improve their L2 performance when repeating the same or similar tasks (Qiu & Lo, 2017).

Furthermore, the components of our debate task design—which range from reading articles and summarizing them to writing cases, engaging in actual debates, and processing the instructor's feedback on written cases—form a coherent whole which smooths the path for the transfer of gains. This would be expected to enable students, for example, to employ some of the words they learn from articles in the cases they subsequently write. Earlier research documented transfer patterns in tasks whose activities are closely related (Jianling, 2018) and tasks that have a generic relationship (Hyland, 2007; So, 2005). In our

TABLE 4
Frequency of the Use of Sophisticated Connectives

	Control ($n = 51$)		Intervention ($n = 95$)	
Measure	Pretest	Posttest	Pretest	Posttest
Moreover	0	2	1	32
Furthermore	1	2	4	44
Besides	5	3	0	16
In addition	0	2	1	7
Additionally	0	1	0	4
However	0	1	4	7
Therefore	1	4	2	12

intervention, the preparatory articles that the intervention students read overlapped quite significantly with the written texts produced in terms of their genre and communicative purpose (convincing someone of a particular standpoint).

Now we will look at our findings in more detail.

Fluency

The participants in the intervention group showed significant improvement in their writing fluency (measured via text length) compared to the control group. Though the control group also practiced writing, the intervention group, quantitatively speaking, practiced slightly more, and this may have impacted their writing fluency. However, there are many relevant intervention-related factors that have arguably contributed to this significant improvement. Writing in the debate environment is a socially oriented activity and provides learners with a meaningful and functional purpose, as well as an authentic audience. Students write cases about topics that are relevant to their interests and which they can relate to and deliver to an authentic audience. Research has demonstrated that such a learning environment motivates students and enhances their writing fluency (e.g., Albadi, 2016). In addition, it seems that the repeated purposeful practice of writing in the debate intervention may have facilitated the process of automatization of at least a number of formulaic expressions and cohesive devices. As a result, the debaters might have developed a degree of automaticity in these expressions (e.g., *I am in favor of the death penalty because of the following reasons…*; see DeKeyser, 2007).

Note also that the learners were instructed to read at least two argumentative articles prior to each debate. It seems that the learners profited from the recurrent process of reading and writing that enabled them to leverage these resources better and better, and hence enhanced their writing fluency. Earlier research has demonstrated the effectiveness of L2 reading-to-write pedagogy, finding that extensive reading helps students to improve their vocabulary (lexical complexity) and to write better (e.g., Hirvela, 2016; Hyland, 2019), as "extensive reading can furnish a great deal of both tacit and conscious knowledge of conventional features of written texts, including grammar, vocabulary, organizational patterns, interactional devices and so on" (Hyland, 2019, p. 17).

Accuracy

The debate-based instruction also appears to offer some advantages in terms of promoting learners' writing accuracy. We found that the intervention group made more improvements than the control group in four (out of five) indices of accuracy, with three measures reaching statistical significance (EFCs, lexical errors/100 and morphological errors/100). The control group also received feedback on their writings, which may have led to some improvement (see Table 2). However, the feedback environment in debate seems to be more effective. At the end of each debate, the participants in the intervention group submitted their written cases, on which they received feedback from the instructor. Some students revised and resubmitted their work the following week. Additionally, during debates, the participants were instructed to note the mistakes their classmates made and improve them. Moreover, the instructors occasionally discussed some of the commonly made mistakes during actual debates. Through this rich feedback environment, the learners had the opportunity to monitor their progress, recognize gaps in their output, and reflect on their writing performance. The recurrent process of crafting a case prior to each debate and receiving feedback on it seems to have enabled the debaters to apply the knowledge they gained from feedback on a particular case when writing a new one.

It is worth noting that speech delivery and the exchange of arguments in front of classmates may have put some pressure on students to try their best not to make embarrassing grammatical mistakes. Also, the debate environment seems to raise awareness of the importance of grammatically accurate language in conferring cogency on arguments. Past research has shown that when learners' writing is aimed at an authentic audience, they tend to be more precise and accurate (Albadi, 2016) and "to craft higher quality and more sophisticated arguments" (Chen, Hand et al., 2016, p. 308).

Syntactic Complexity

Results indicate that the debate intervention resulted in improving one measure of syntactic complexity, MLC, which taps complexity at the clausal and phrasal level (Pallotti, 2015). Research on syntactic complexity has identified this index as one of the strongest measures to capture syntactic complexity development (e.g., Lu, 2010; Norris & Ortega, 2009). MLT and CT measures did not reach significance. The mean of MLT produced at pretest and posttest for both groups did not change greatly. This finding is in line with the findings of previous studies (e.g., Knoch,

Rouhshad, & Storch, 2014; Ortega, 2003; Storch, 2009). Ortega (2003), for example, speculated that it takes 12 months of language instruction for MLT to improve substantially. As for subordination (i.e., C/T), while its mean remained almost the same for the control group (pretest: $M = 1.85$; posttest: $M = 1.83$), it slightly decreased for the intervention group at posttest (pretest: $M = 1.84$; posttest: $M = 1.76$). Some scholars have associated subordination decrease with L2 proficiency increase as other forms of complexity (e.g., phrasal complexification) increase (Norris & Ortega, 2009).

Lexical Complexity

The intervention group gained the most improvement in all measures of lexical sophistication, with WL reaching significance. Both WF and WL were found to be effective predictive indices of essay writing quality (McNamara et al., 2010). Also, WL and WF have been regarded as indicators of lexical proficiency, as learners who produce less frequent words (Crossley, Salsbury, & McNamara, 2014) and longer words (Yoon, 2017) are judged to be more lexically proficient.

One interpretation of the positive effect of in-class debates on lexical complexity might be attributable to the lexically rich environment in which the debates engaged the students. Debate seems to be fertile ground for the acquisition of new vocabulary and its active implementation. Prior to each debate, the participants were asked to read at least two articles. The participants were instructed to read these carefully in order to find relevant arguments. To do this, it was essential to understand, if not all, at least the majority of the vocabulary. Such a process led to deciphering the unknown words in a context relevant to them. Importantly, case writing created an opportunity to use the newly learned words and hence increased their retention. Furthermore, the debate environment invited the implementation of these words, a fact that further reinforced the students' grip on these lexical gains. The contribution of debate to lexical enhancement is consistent with past research (e.g., el Majidi et al., 2018) in which the debaters reported improvement to their vocabulary after participating in L2 debates.

Cohesion

Beyond the well-known areas of fluency, accuracy, and lexical and syntactic complexity, this research documents another performance area that is critical to persuasive texts—namely, the use of a specific set of organizational markers that explicitly signal conceptual relationships between clauses and text fragments. Well-structured texts and coherent arguments are important distinguishing traits of 'good academic writing' (Storch, 2009).

Results revealed improvement in the majority of the assessed cohesive devices in the intervention students' texts. One interpretation of these results might be attributable to the fact that the students in the intervention group received a list of common cohesive devices and were asked to use them in their written cases to establish cohesion between the adduced arguments.[5] More importantly, the debate context seems to provide a natural and fertile ground for the development of this important area of writing proficiency. Debate-case writing prompts students to generate a logical and coherent text by interrelating arguments in a coherent manner and organizing them in a stepwise hierarchical format. For example, presenting different arguments requires the use of transition and frame markers to mark progression and a shift from one argument to another. Through recurrent practice, we believe that the debate environment instilled in the debaters the need to make explicit relations between arguments if they were to persuade and outsmart their opponents. Research revealed that embedding transitional markers in engaging, interesting, and challenging content stimulates their development (Crosson & Lesaux, 2013).

The growth in the use of transitional markers in terms of quality and sophistication further makes a case for the impact of in-class debates on writing proficiency. Past research has demonstrated that cohesive markers, in particular their diversity, are predictive of the quality of argumentative writing (e.g., Qin & Uccelli, 2016; Uccelli et al., 2013). Research has also correlated the use of sophisticated markers with higher quality writing (e.g., Ho & Li, 2018).

CONCLUSION AND LIMITATIONS

Our current findings offer a modest but promising step forward in unravelling the potential of in-class debates for improving L2 writing proficiency. The results revealed that the students in the intervention group tended to write significantly longer texts that exhibited more syntactic complexity in terms of phrasal and clausal complexity, included more sophisticated vocabulary, demonstrated better grammatical accuracy, and contained a more sophisticated and wider range of indices of cohesion after the intervention than

their counterparts in the control group. The significant increase in the majority of the measures employed after the intervention indicates that the debate environment can be a privileged site for developing various linguistic aspects of writing.

This study provides insights into some of the pedagogical merits of the debate environment that can affect L2 writing development. The debate environment (or, at least, our debate intervention) allows learners to learn language and writing through the medium of writing about purposeful content of relevance to them. In other words, it enables the interface and synergy of the three writing perspectives of LW, WLC, and WLL (see Manchón, 2011; Ortega, 2011). Moreover, it promotes an effective interplay between speaking and writing in a way that scaffolds writing development (Yang, 2008) and it lends itself readily to reading-to-write pedagogy, which empowers learners to write better (Hirvela, 2016). Additionally, it allows an iterative cyclic process of writing, feedback processing, and rewriting as well as purposeful and meaningful practice that facilitates attending to linguistic processes that enhance L2 writing development. Furthermore, the debate environment in the intervention seems to stimulate the development of students' metacognitive awareness of processes that lead to writing development. For example, this environment seemingly instills in learners the awareness that the use of sophisticated, coherent, and accurate language enhances the persuasiveness of their discourse. Finally, learners' positive attitude toward debating, its competitive atmosphere, and the presence of a real audience (i.e., teacher and classmates/opponents) seem to provide learners with extra stimuli to pay more attention to different dimensions of language use and content (argumentation).

The findings of this study break with the common view that debates are only suitable for extracurricular activities or in the competitive sphere. We hope that the potential for in-class debates to enhance written performance in L2 teaching contexts will be recognized and that L2 instructors will consider employing them on a regular basis in their teaching practice. As it may not be feasible to plan debates on a weekly basis, we recommend scheduling them biweekly or once every 3 weeks. Debates are favored by students and hold potential for honing other skills. In other words, there is more than one reason to engage students in debates on a regular basis.

To the best of our knowledge, this is the first study that has empirically investigated the impact of in-class debates on writing proficiency in the L2 or FL context. As such, it lays the groundwork for future work that may further explore the potential of these findings and their implications for L2 writing teaching and learning. Therefore, the contribution of the present study should find continuation in further research.

While promising, the study must be viewed in light of some important limitations that might preclude the wide generalizability of our findings. Although significant results were found, our sample is not representative of all secondary-school students. Analysis of additional writing samples from a larger variety of schools and students would be necessary to accurately measure the effects of debate-based instruction on writing proficiency and confirm these findings.

An additional note of caution is required: This study assessed students' written performance in a task that may have slightly favored the intervention students over their control counterparts. Though the control group also practiced writing tasks that were argumentative in nature (writing persuasive essays and letters), the intervention students were engaged in a pedagogical environment that more strongly supports the expression of opinion. Since writing performance tends to vary significantly across task types and genres, future studies should consider assessing proficiency using a variety of different and independent writing tasks to test to what extent the effects are transferable. To assess carryover effects, we have gathered additional writing samples (essays and letters) from both groups, which we will analyze in a future study with the same genre-independent measures used in this study.

Last, this study elicited data from two time points (pre- and posttest). Future studies would also benefit from a delayed posttest time point. This third time point would yield insight into the durability of the potential effects and hence provide additional evidence for the performance patterns observed.

NOTES

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

[3] Dobbs used persuasive tasks to represent opinion tasks.

[4] Unlike Yoon and Polio (2017), we included lexical errors, which we defined as errors in word choice or

word form. As for prepositional errors, we included errors in all types of constituents.

[5] It is important to note that the students were not given instructions about different sorts of transitional markers. Also, note that the students were not asked to use these connectors during pretest and posttest opinion tasks.

ACKNOWLEDGMENTS

This work has been financed by the Netherlands Organization for Scientific Research (NOW; grant number 023.009.059).

REFERENCES

Aclan, E. M., & Aziz, N. H. A. (2015). Exploring parliamentary debate as a pedagogical tool to develop English communication skills in EFL/ESL classrooms. *International Journal of Applied Linguistics and English Literature*, *4*, 1–16.

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.

Allen, H. W. (2018). Redefining writing in the foreign language curriculum: Toward a design approach. *Foreign Language Annals*, *51*, 513–532.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (release 2)* [CD]. Philadelphia, PA: University of Pennsylvania, Linguistic Data Consortium.

Badger, R., & White, G. (2000). A process genre approach to teaching writing. *ELT Journal*, *54*, 153–160.

Beeker, A., Fasoglio, D., Jong, K. D., Keuning, J., & Til, A. V. (2015). ERK-niveau schrijfvaardigheid Engels, Duits en Frans; Onderzoek naar het bereikte niveau aan het eind van havo en vwo [CEFR-level of writing skill in English, German and French; Research into the achieved level at the end of *havo* and *vwo*]. *Levende Talen Tijdschrift*, *16*, 16–25.

Belcher, D., & Hirvela, A. (Eds.). (2008). *The oral-literate connection: Perspectives on L2 speaking, writing, and other media interactions*. Ann Arbor, MI: The University of Michigan Press.

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*, 277–320.

Chen, Y. C., Park, S., & Hand, B. (2016). Examining the use of talk and writing for students' development of scientific conceptual knowledge through constructing and critiquing arguments. *Cognition and Instruction*, *34*, 100–147.

Cho, H. (2017). Synchronous web-based collaborative writing: Factors mediating interaction among second-language writers. *Journal of Second Language Writing*, *36*, 37–51.

Council of Europe. (2018). The common European framework of reference for languages: Learning, teaching, assessment—Companion volume with new descriptors. Strasbourg, France: Author. Accessed 27 June 2020 at https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989

Crossley, S. A., Salsbury, T., & Mcnamara, D. S. (2014). Assessing lexical proficiency using analytic ratings. *Applied Linguistics*, *36*, 570–590.

Crosson, A. C., & Lesaux, N. K. (2013). Connectives. *The Reading Teacher*, *67*, 193–200.

DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 97–113). New York: Routledge.

Dewaele, J. M., Witney, J., Saito, K., & Dewaele, L. (2018). Foreign language enjoyment and anxiety: The effect of teacher and learner variables. *Language Teaching Research*, *22*, 679–697.

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. (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*, 924–932.

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.

Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing*, *10*, 161–184.

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

Hirvela, A. (2016). *Connecting reading and writing in second language writing instruction* (2nd ed.). Ann Arbor, MI: The University of Michigan Press.

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

Ho, V., & Li, C. (2018). The use of metadiscourse and persuasion: An analysis of first year university students' timed argumentative essays. *Journal of English for Academic Purposes*, *33*, 53–68.

Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. London: Continuum.

Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing*, *16*, 148–164.

Hyland, K. (2011). Learning to write: Issues in theory, research, and pedagogy. In R. M. Manchón (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 17–35). Amsterdam: John Benjamins.

Hyland, K. (2019). *Second language writing* (2nd ed.). Cambridge: Cambridge University Press.

Inoue, C. (2016). A comparative study of the variables used to measure syntactic complexity and accuracy in task-based research. *The Language Learning Journal*, *44*, 487–505.

Jianling, L. (2018). The impact of face-to-face oral discussion and online text-chat on L2 Chinese writing. *Journal of Second Language Writing*, *41*, 27–40.

Knoch, U., Rouhshad, A., & Storch, N. (2014). Does the writing of undergraduate ESL students develop after one year of study in an English-medium university? *Assessing Writing*, *21*, 1–17.

Leki, I., Cumming, A., & Silva, T. (2010). *A synthesis of research on second language writing in English*. New York: Routledge.

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

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.

Manchón, R. M. (2011). Situating the learning-to-write and writing-to-learn dimensions of L2 writing. In R. M. Manchón (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 3–14). Amsterdam: John Benjamins.

Manchón, R. M., & Williams, J. (2016). L2 writing and SLA studies. In R. M. Manchón & P. K. Matsuda (Eds.), *The handbook of second and foreign language writing* (pp. 567–586). Boston: de Gruyter.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, *27*, 57–86.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, *20*, 83–108.

Miller, J. (2008). *An introduction to English syntax* (2nd ed.). Edinburgh, UK: Edinburgh University Press.

Naghdipour, B. (2016). English writing instruction in Iran: Implications for second language writing curriculum and pedagogy. *Journal of Second Language Writing*, *32*, 81–87.

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.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, *24*, 492–518.

Ortega, L. (2011). Reflections on the learning-to-write and writing-to-learn dimensions of second language writing. In R. M. Manchón (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 237–250). Amsterdam: John Benjamins.

Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research*, *31*, 117–134.

Pica, T., Kang, H. S., & Sauro, S. (2006). Information gap tasks: Their multiple roles and contributions to interaction research methodology. *Studies in Second Language Acquisition*, *28*, 301–338.

Plakans, L., Gebril, A., & Bilki, Z. (2016). Shaping a score: Complexity, accuracy, and fluency in integrated writing performances. *Language Testing*, *36*, 161–179.

Polio, C., & Park, J. H. (2016). Language development in second language writing. In R. M. Manchón & P. K. Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 287–306). New York: Routledge.

Qin, J., & Karabacak, E. (2010). The analysis of Toulmin elements in Chinese EFL university argumentative writing. *System*, *38*, 444–456.

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.

Qiu, X., & Lo, Y. (2017). Content familiarity, task repetition and Chinese EFL learners' engagement in second language use. *Language Teaching Research*, *21*, 681–698.

Sasaki, M. (2004). A multiple-data analysis of the 3.5-year development of EFL student writers. *Language Learning*, *54*, 525–582.

Schmidt, R. W. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge: Cambridge University Press.

Shum, M. S. K., Tai, C. P., & Shi, D. (2018). Using 'reading to learn' (R2L) pedagogy to teach discussion genre to non-Chinese-speaking students in Hong Kong. *International Journal of Bilingual Education and Bilingualism*, *21*, 237–247.

Snider, A., & Schnurer, M. (2006). *Many sides: Debate across the curriculum*. New York: International Debate Education Association.

So, B. P. (2005). From analysis to pedagogic applications: Using newspaper genres to write school genres. *Journal of English for Academic Purposes*, *4*, 67–82.

Stewart, T. (2003). Debate for ESOL students. *TESOL Journal*, *12*, 9–15.

Storch, N. (2009). The impact of studying in a second language (L2) medium university on the development of L2 writing. *Journal of Second Language Writing*, *18*, 103–118.

Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *The Canadian Modern Language Review*, *50*, 158–164.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. *Planning and Task Performance in a Second Language*, *11*, 239–273.

Turgut, Y. (2009). EFL learners' experience of online writing by PBWiki. In G. Siemens & C. Fulford

(Eds.), *ED-MEDIA 2009–World conference on educational multimedia, hypermedia, & telecommunications* (pp. 3838–3847). Waynesville, NC: AACE.

Uccelli, P., Dobbs, C. L., & Scott, J. (2013). Mastering academic language: Organization and stance in the persuasive writing of high school students. *Written Communication, 30*, 36–62.

Wade, M. M. (1998). The case for urban debate league. *Contemporary Argumentation and Debate, 19*, 60–65.

Williams, J. (2012). The potential role(s) of writing in second language development. *Journal of Second Language Writing, 21*, 321–331.

Wolfe–Quintero, K., Inagaki, S., & Kim, H. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu, HI: University of Hawai'i at Manoa, Second Language Teaching and Curriculum Center.

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). Ann Arbor, MI: The University of Michigan Press.

Yang, Y. (2016). Teaching Chinese college ESL writing: A genre-based approach. *English Language Teaching, 9*, 36–44.

Yong, F. L. (2010). Attitudes toward academic writing of foundation students at an Australian-based university in Sarawak. *European Journal of Social Sciences, 13*, 471–477.

Yoon, H. J. (2017). Linguistic complexity in L2 writing revisited: Issues of topic, proficiency, and construct multidimensionality. *System, 66*, 130–141.

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.

Zare, P., & Othman, M. (2013). Classroom debate as a systematic teaching/learning approach. *World Applied Science Journal, 28*, 1506–1513.

APPENDIX

An Example of the Analysis of a Text Produced by a Participant

I am in favor of the death penalty. and I have a few argument[mor] to support my side. First of all, the death penalty is a good way to scare criminals off and makes them think twice before they commit a crime. Also, in my opinion, criminals don't deserve to stay alive if they took another person's life or did something terrible to another human being. The people who have been murdered were not given a second chance. So why should the[mor] murderers get one. Furthermore, as cruel as it may sound, making the death penalty a form of punishment again will save us a lot of money. We wouldn't have to pay for cells, food and facilities for the[mor] prisoners anymore, which will make our taxes much lower. Lastly, the death penalty will make grieving easier for family and friends of victims of murder[mor] or for the victims of rape or robbery. Knowing that the person who did something to you or a loved one, won't walk around on the streets anymore will make you feel a lot safer.

Note: mor = morphological error

Measure	Index	Value
Fluency	Number of words	184
Syntactic complexity	MLT	18.40
	MLC	1.80
	C/T	10.22
Lexical complexity	MLTD	95
	Word frequency	2.99
	Word length	4.14
Accuracy	EFCs	0.84
	Lexical errors[a]	0
	Syntactic errors[a]	0
	Morphological errors[a]	2.17
	Preposition errors[a]	0
Cohesion	Transition markers	3
	Frame markers	2
	Gloss markers	0
	Conclusion markers	0
	Diversity of type	2
	Diversity of token	5

Note. 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; EFCs = error-free clauses.
[a]Per 100 words.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.