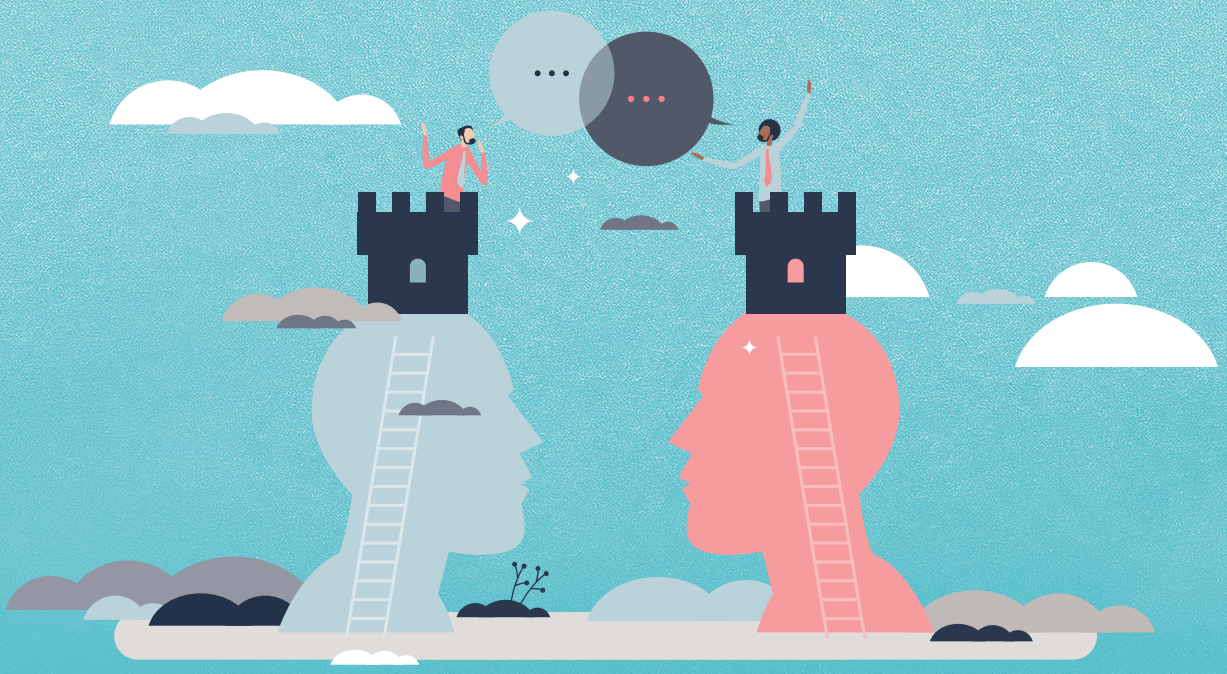
Debate as a Tool for L2 Learning

Investigating the Potential of In-Class Debates for Second Language Learning and Argumentation Skills



Abid El Majidi

Debate as a Tool for L2 Learning

Investigating the Potential of In-Class Debates
for Second Language Learning
and Argumentation Skills

Abid El Majidi

Printing: Ridderprint | www.ridderprint.nl

ISBN/EAN: 978-94-6458-190-4

Debate as a Tool for L2 Learning

Investigating the Potential of In-Class Debates for Second Language Learning and Argumentation Skills

Debat als instrument voor tweedetaalleren

Onderzoek naar de waarde van debatteren in de klas voor het leren van een tweede taal en argumentatievaardigheden

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

vrijdag 20 mei 2022 des ochtends te 10.15 uur

door

Abid El Majidi

geboren op 20 maart 1980
te Driouch, Marokko

Promotor:

Prof. dr. H.C.J. de Graaff

Copromotor:

Dr. D.M.L. Janssen

Assessment Committee:

Prof. dr. H.H. van den Bergh
Dr. N.H. de Jong
Prof. dr. M.C. Michel
Prof. dr. A. Revesz
Prof. dr. T.J.M. Sanders

Dit proefschrift werd (mede) mogelijk gemaakt met financiële steun van de NWO Promotiebeurs voor Leraren, projectnummer: 023.009.059

ico

Interuniversity Centre for Educational Sciences

Table of Contents

General Abstract

Debate is widely hailed as an effective teaching tool capable of attaining many learning objectives. Yet, in second language education there is scant empirical evidence available to support this assumption. The present study set out to: 1) elicit L2 students' perceptions with regard to the factors that underlie positive attitudes towards debate and their perceptions of its impact on different aspects of language development; and 2) experimentally explore the affordances and effectiveness of debate pedagogy with regard to speaking, writing, and argumentation skills in the Dutch secondary school context.

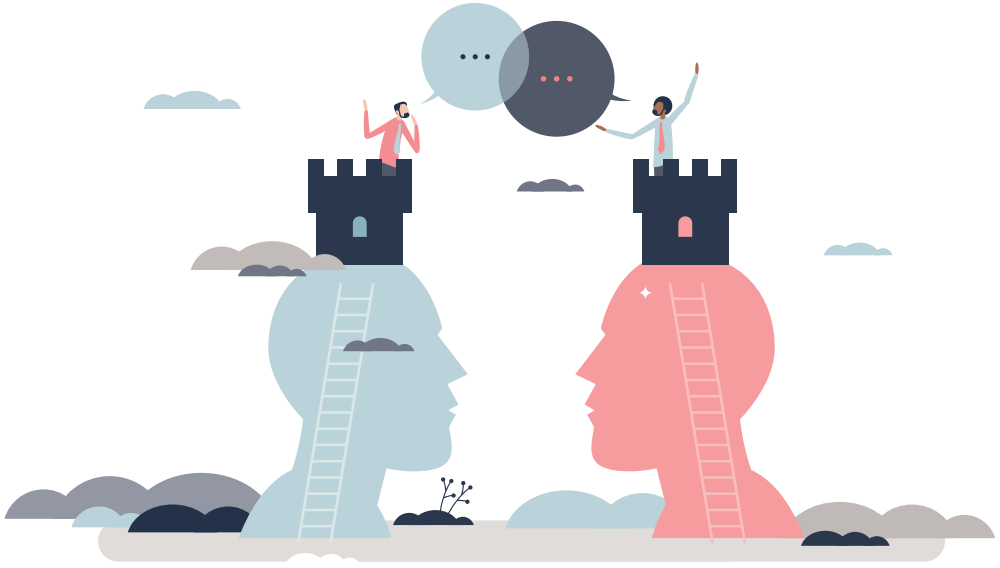
After having developed a debate task design—consisting of a number of speaking and writing activities—following the guidelines and principles of educational design research, we conducted the perception studies. Drawing on data from questionnaires, interviews, classroom observations, and student performance, the main finding of the perception studies was that the participants greatly admire debate and correlate debate participation with an improvement in all language areas being investigated (i.e., speaking, writing, reading, and listening skills, as well as vocabulary and grammar).

The intervention studies, which were conducted in three secondary schools and involved intervention and control groups, evaluated the effects of debate pedagogy on different aspects of speaking, writing (including opinion writing tasks and argumentative essays), and argumentation skills. Multilevel analyses in all these intervention studies revealed that the students in the intervention group significantly improved on a number of the assessed measures in comparison to the control group.

From the findings, it can be concluded that debate pedagogy can provide a motivating learning environment with a unique pedagogical configuration that is conducive to L2 language and argumentation performance. The practical implications of the findings as well as avenues for future research are addressed.

Chapter 1

Introduction



"The use of debate as a teaching technique in the classroom can fulfil several major objectives … [such as] problem solving, critical thinking, oral and written communication, and research can be approached. Debate is not intended to be an end in itself, but, rather, a means to desirable educational outcomes." (Jackson, 1973 pp. 152–153)

Abstract

This chapter presents the context, concepts, rationale, and motivation for the current PhD thesis. The chapter starts by touching upon the interplay between pedagogy and second language acquisition (SLA). Next, the challenges of learning and teaching a second language (L2) are briefly discussed, followed by my personal inspiration for, and the aim of, this research. The chapter concludes with an outline of the dissertation and some concluding remarks.

1.1 L2 pedagogy and SLA

There is a long-standing debate about the relationship between SLA research and L2 pedagogy. Some argue that although SLA research has made much progress, its influence on, and utility to, L2 teaching has been very limited (Block, 2000; Ellis, 2021; Klein, 1998; Stapleton, 2014). Klein (1998) said that "over the last 25 years, second language acquisition research has made considerable progress in more than one respect. Still, we are very far from proving a solid basis to foreign language teaching" (p. 572). Ellis (1997b) attributed the challenge of connecting SLA research to L2 pedagogy to the fact that:

> researchers have been busy with finding out how L2 learners acquire a second language, while teachers have been busy trying to help them do it. Researchers have been wary of making proposals based on their research. Teachers have not had the time (or perhaps the inclination) to find out what researchers have discovered. (p. 3)

To bridge this gap between SLA and L2 pedagogy, Ellis (1997a) argued that an educational perspective is needed. This perspective should suggest "ways in which SLA can be appraised in a pedagogically relevant manner and, more importantly, what kinds of applications may be fruitful" (p. 69).

One kind of research whose relevance to L2 pedagogy is invaluable is research conducted by teachers themselves (Ellis, 2010; Loewen, 2020; Nassaji, 2012; Sato & Loewen, 2019a, 2019b). This kind of research is relevant, in that teachers—as assessors in this case—are familiar with the practice of classroom teaching. So "to be successful, researchers need teachers' help, knowledge, and participation in making valid decisions about the connection between research and practice" (Nassaji, 2012, p. 342). Lightbown (2000) maintained that SLA could not function as a source of telling teachers what and how to teach and that it is up to teachers themselves to decide what is pedagogically relevant and what is not. In the same vein, Stewart (2006) has called for "teachers' research," that is research conducted by teachers for teachers. Sato and Loewen (2019b) argued that "teachers themselves are most knowledgeable about their own L2 teaching/learning issues. Additionally, teachers' experiences both as L2 learners and teachers are invaluable sources for improving pedagogy" (p. 9). This type of research has been equally regarded as a means of enhancing professional development (Crookes & Chandler, 2001), fostering reflective teaching, and discovering alternative perspectives on pedagogical problems (Wallace, 1998).

This kind of research has been recently encouraged and received increasing attention in the Netherlands. The Dutch Ministry of Education, in collaboration with the Dutch Research Council (NWO), has initiated doctoral grants for teachers, including teachers in primary and

secondary education. The primary aim of this initiative is to directly apply teachers' expertise, knowledge, and research skills acquired for the benefit of educational practice and hence improve the quality of education. This program is likewise expected to strengthen the ties between universities and schools. The present PhD study has been funded by this initiative.

1.2 Learning and teaching an L2

Learning an L2 is an arduous task, as it involves a nexus of interwoven factors and is hedged around with many caveats, such as the scarcity of opportunities to use it outside the classroom. Likewise, teaching an L2 is a challenging task that requires dynamic, creative, and flexible ways to manage the learning process. In the face of these challenges, L2 instructors carry the responsibility of seeking out innovative pedagogical tools that involve learners in motivational, engaging, and meaningful tasks that bring about learning most effectively and to a higher standard.

L2 learners arrive in L2 classrooms with the expectation of achieving a particular level in order to fare well in academic and professional careers. Yet, many teachers and researchers alike have indicated that many L2 learners do not achieve an acceptable level of competency that allows them to operate effectively, fluently, and confidently in the target language (Piggott, 2019). This seemingly and unsurprisingly leads to frustration and demotivation.

1.3 The inspiration for this research

The initial motivation that drove this research came from my own experience as a teacher of English in a Dutch secondary school. I have used debates for many years in my teaching practice. I have observed with a sense of gratification the high rate of interest and motivation among my students when engaged in debating. Students have repeatedly asked me to plan debates on a regular basis in our classes, and they have always fully participated with enthusiasm in the learning experience.

I initially turned to debate to practice speaking skills. Over the years, I was pleasantly surprised by the growth in students' confidence in their speaking ability and in their ability to think fast and critically about complex issues. It was truly remarkable to observe how—in my experience—debate develops and reinforces many skills that align well with different learning objectives. Debate seemed to be a valuable teaching method that holds the potential to create a rigorous learning environment. There were, however, no empirical studies that objectively tested the effects of debating on the language proficiency and argumentation skills of L2 students. This research journey, therefore, sets out to investigate this pedagogical potential of debate, which could have implications for L2 pedagogy.

1.4 The aim of the study

Given the current situation of L2 teaching, which is surrounded by many challenges entailing the search for pedagogical innovative tools, the present study was conducted in order to expand the research-based knowledge about effective L2 pedagogy. Improving L2 pedagogy requires investigating "the effects of instruction on learning empirically. Such empirical research [provides] a means of testing the claims of specific theoretical positions and it might contribute to better, research-based pedagogy" (Ellis & Shintani, 2014, p. 16).

The aim of the current research is to focus on the use of debate pedagogy as a way to offer a potential remedy to concerns relating to teaching and learning L2 in secondary schools (see Chapter 3). The effectiveness of debate pedagogy was tested in intervention studies conducted in three secondary schools.

Many studies have suggested that debates can operate as an effective mechanism for fulfilling many learning objectives in the L2 teaching context. Debates engage students in a rich and authentic learning environment that entails researching, synthesizing, discussing, and analyzing inspiring topics, a fact that gives students a meaningful and rewarding learning experience (Fallahi & Haney, 2007; Omelicheva & Avdeyeva, 2008; Tumposky, 2004). Debates integrate skills in an interesting way and create a healthy competitive atmosphere that induces meaningful and rich interactions (Stewart, 2003). What is more, students' enjoyment of debates reinforces the richness and effectiveness of the learning process (e.g., Tessier, 2009) and presumably of learning outcomes.

Debate "is not an end in itself but, rather, a task through which language practice can be orchestrated" (Stewart, 2003, p. 15). It is hailed as a conducive platform for reaping many benefits in an L2 context, particularly improving language skills (e.g., Lustigova, 2011; Zare & Othman, 2013) and promoting "argumentation skills for persuasive speech and writing" (Krieger, 2005, p. 25).

There has been little research to date that has explored the pedagogical affordances of the debate environment for L2 learners, especially when it comes to "the impact of debate on middle school students, a population largely absent from the literature" (Mirra et al., 2016, p. 5) and "characterized by low language proficiency levels, and few opportunities to engage with the FL[/L2]" (Villarreal & Gil-Sarratea, 2020, p. 874). The debate research literature is characterized by the dominance of anecdotal research (e.g., perceived effects) that has highlighted the potential of debate for enhancing language skills and argumentation skills. However, as Niu (2009) stated:

> learning potentials do not equal actual learning, because learning potentials are involved in the process of learners' task performance, while actual learning refers to the learning outcome, that is, whether the learners can recall or use the related linguistic knowledge correctly some time after task treatment. In fact, actual learning gains can be measured by using pretests and posttests. (p. 398)

To the best of my knowledge, no study to date has experimentally examined the effects of debate pedagogy on writing and speaking skills across all main dimensions of language production (fluency, complexity, accuracy, adequacy, and cohesion) and argumentation skills (structural patterns and quality aspects of argumentation). To accurately gauge the effects of debate pedagogy on these skills, experimental research is needed (Kennedy, 2009; O'Mahoney, 2015; Tumposky, 2004). Only with empirical effect studies are we able to track the claimed and perceived benefits of debate on language and rhetorical development and hence establish their validity (Inoue & Nakano, 2004; Ortega & Iberri-Shea, 2005; Randolph, 2007). To safeguard the reliability and validity of the findings establishing a link between debate pedagogy and effects on language proficiency and argumentation skills, many scholars called for the use of a pretest–posttest design with a control group (Camp & Schnader, 2010; Oros, 2007; Randolph, 2007).

1.5 Overview of the dissertation

The main objective of this dissertation is to provide insights into the impact of in-class debates on different dimensions of the speaking, writing, and argumentation skills of Dutch secondary students. To meet this end, the thesis includes 10 chapters. After the introduction chapter, which has set the scene for the current research, the dissertation proceeds as follows. Chapter 2 provides a review of the state of the art of research relevant to this study with a focus on research studies highlighting some gaps leading to the current research. This chapter also reviews how debates are implemented in teaching practice with the aim of identifying instructional guidance that can inform the construction of a debate task design.

As preparation for the intervention studies that I conduct in the next chapters, Chapter 3 seeks to engineer a comprehensive and workable debate task design premised on pedagogically sound ground that can ensure optimal preparation and participation of all students. More specifically, this chapter examines task design principles and factors that can lead to effective debates, accommodating the particularities of the Dutch secondary school context.

Debate is widely appreciated by students. However, we know little about students' attitude towards debate in the L2 secondary school context. Task attitude is an impactful variable that can affect students' involvement and performance (Dewaele et al., 2018). Chapter 4, therefore, investigates the extent to which L2 Dutch secondary school students hold a positive attitude

towards debate and explores the factors that underlie this attitude towards this instructional tool.

Chapter 5 elicits the participants' perceptions and perspectives regarding the effectiveness of in-class debates for developing their English language proficiency across the skill areas of speaking, writing, reading, and listening, as well as vocabulary and grammar. Most of the existing debate-perception studies so far have involved university students. This study also investigates the relationship between attitude towards debate and the perceived effects as well as performance.

Chapters 6–9 present the findings of the debate intervention studies—conducted in three secondary schools, involving intervention and control groups—on different aspects of speaking, writing, and argumentation skills. To the best of my knowledge, no empirical studies have investigated the effects of debate pedagogy on different dimensions of these skills. Having empirical-based evidence is likely to provide a more reliable picture of the impact of debate pedagogy (on the assessed performance areas) and may also motivate instructors to employ debates in their teaching practice.

For these chapters (6–9), measuring L2 performance and development has been operationalized following a multidimensional approach that includes measures—identified by the relevant literature as key indicators for L2 performance and growth—for complexity, accuracy, fluency, adequacy, and cohesion (CAFAC). In this study, communicative adequacy (also known as functional adequacy) has been operationalized in terms of the assessment of different aspects of argumentation, as effective argumentation reflects the effectiveness of task success in this study (i.e., convincing someone of a particular point of view). In addition to representing the adequacy construct, argumentation is concurrently studied in its own right (Chapter 8) to gain insights into the extent to which debate pedagogy may affect the argumentative competence of debaters.

Chapter 6 investigates the effects of debate pedagogy on different dimensions of L2 speaking competence. The study analyzes speech samples elicited through opinion tasks on two occasions (pre- and posttest) for fluency, complexity, accuracy, and cohesion.

In Chapter 7 the focus shifts from oral to written language production. This chapter examines the effects of the debate intervention on different aspects of writing competence, including fluency, complexity, accuracy, and cohesion. To measure the effects of the intervention, two free opinion tasks (as pretest and posttest) are compared.

The study presented in Chapter 8 discusses the effects that debate pedagogy may have on the argumentative competence of debaters. Chapter 8, therefore, discusses the effects of the debate intervention on a number of L2 structural and quality aspects of L2 argumentation skills

by analyzing written and oral opinion tasks produced by intervention and control groups at pretest and posttest.

Chapter 9 explores the potential of debate pedagogy for improving L2 argumentative essay writing. This study, which employs a pretest–posttest–delayed posttest design with a control group, analyzes argumentative essays for different linguistic and content features, including fluency, complexity, accuracy, cohesion, and communicative adequacy.

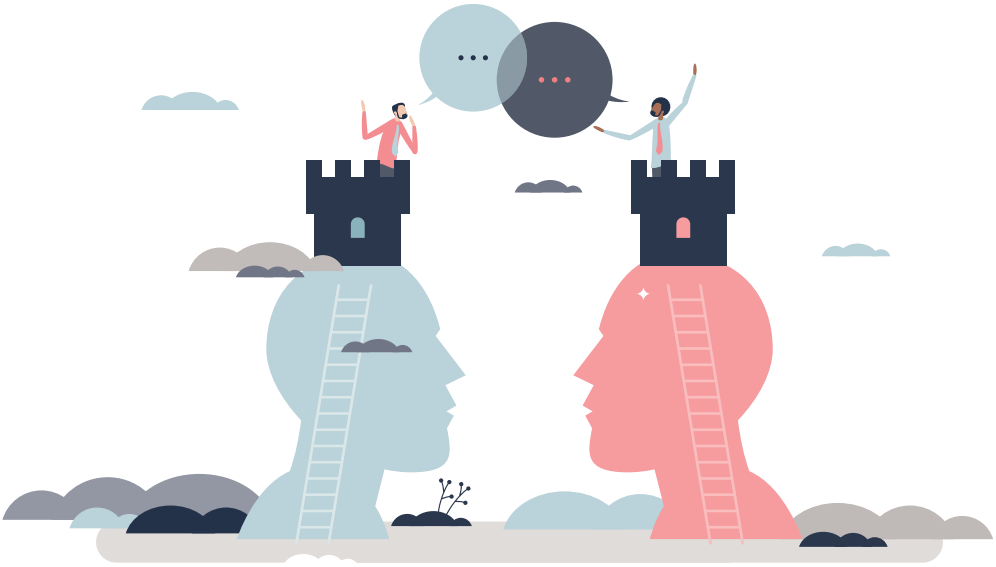
The final chapter, Chapter 10, brings the various strands of the dissertation together. It summarizes the main findings and discusses the pedagogical mechanisms and affordances of the debate environment. This chapter further includes practical implications for teaching practice, discussion of the limitations of the research, and possible directions for future research.

1.6 Final remarks

Several chapters in this dissertation have been written as separate articles. The studies in Chapters 4, 5, 6, 7, and 8 were accepted and published in peer-reviewed academic journals. Therefore, the reader may experience some overlap, particularly in the method section. The study in Chapter 9 has been submitted to a peer-reviewed academic journal and is currently under review. Chapters 4 and 7 have also been published in peer-reviewed professional journals, in a shorter form. Each study is self-contained and can therefore be read independently.

Chapter 2

Debate as a Teaching Tool: A Literature Review



"Debate enables students to express their ideas in a safe environment, where they can have fun and experiment with a new language while being challenged by other students. It encourages students to take risks with language and to think and process in the target language." (Claxton, 2008, p. 61)

Abstract

Debate has been widely credited as an effective teaching tool. As we shall see, debate pedagogy holds potential for reaping many benefits. In this chapter, I discuss a number of these benefits and review how debates are staged in teaching practice. To this end, I conducted a literature research whose aim was twofold: (1) to highlight research gaps that underlie this study; and (2) to identify base knowledge that informs the construction of a workable and comprehensive debate task design that will be addressed in the next chapter.

The chapter is structured as follows. First, I present the working definition of debate that is used in this dissertation. Next, I discuss a number of benefits that debate confers on debaters. I will especially focus on two benefits—enhancement of language proficiency and argumentation skills—as these two benefits form the basis for this PhD study. Then, I present a number of criticisms leveled against debate, and I will respond to them. In the last section, I provide an overview of how debates are implemented in the first (L1) and second language (L2) classroom.

2.1 Definition of debate

Debate is an inherent part of our life and lies at the core of our interactions. We are regularly involved in persuasive attempts to sway the ideas and decisions of others. It is all around us, on the television, on social media, and at home. Much of the most critical decisions, personal and governmental, that affect our lives are taken after debates. We tend to debate about everything from petty matters to life-changing ones. Moreover, cultural diversity, political divergence, and new global developments have created new issues whose resolution demands debate (Davis et al., 2016; Snider & Schnurer, 2006). Debate is mostly about change. It is in the nature of our existence that we are often not content with the status quo and that we crave to change for the better. Debate offers the operational framework that facilitates the negotiation of this change (Davis et al., 2016).

While many processes can be characterized as "debate," it is important to define what is meant by debate in this research. There is not one single definition of debate universally adopted. Different definitions have been advanced over the years. I will present here three of these definitions that are widely cited and are relevant to this study. Branham (1991) defined debate as "the process by which opinions are advanced, supported, disputed, and defended" (p. 1). This broad definition highlights the main process lying at the core of debating, namely supporting one's side with arguments and questioning the opponent's line of reasoning. Freeley and Steinberg (2005) defined debate as "the process of inquiry and advocacy, a way of arriving at a reasoned judgment on a proposition" (p. 6). This definition points, in addition to the process underlying debate, to the goal of debate, which is "arriving at a reasoned judgment on a proposition." Lastly, Snider and Schnurer (2006) conceptualized debate as "an equitably structured communication event about some topic of interest, with opposing advocates alternating before a decision-making body" (p. 5). This conceptualization seems to be more comprehensive and concurs better with the perspective of this study that classroom debates need to be formalized and structured. In other words, debates should be equitably structured around a topic of interest, and all "sides" should be awarded an equal amount of time to express their views and to expose weaknesses in their opponents' arguments.

Lubetsky et al. (2000) considered debate a highly "sophisticated form of immediate, interactive communication . . . [which] assumes a high level of discourse skill" (p. 10). It involves "a complexity that extends far beyond the level of ordinary conversation, demanding active and critical listening, as well as advanced linguistic competency and critical thinking" (Lieb, 2007, pp. 73–74). Garrett et al. (1996) and Hall (2011) argued that debate differs from discussion, in that debate entails a prearranged position on an issue (for/affirmative or against/negative positions). Additionally, unlike discussions, in-class debates are set up in accordance with

debate formats that go through a number of stages (e.g., constructive speech, rebuttal, etc.). Lastly, while discussions can simultaneously involve the whole class, in-class debates are undertaken in groups, mostly consisting of three to six debaters.

It is worthwhile noting that debate as a teaching tool has been referred to in the literature in different terms: educational debate (e.g., Randolph, 2007), classroom debate (e.g., Scott, 2008), academic debate (e.g., Mitchell, 1998), and in-class debate (e.g., Kennedy, 2007). In this study, I will use the last term, since it explicitly refers to debates that take place in class.

2.2 Debate as a teaching tool

Reports about debates date back over 4,000 years to the Egyptians (2080 B.C.) (Kennedy, 2007). Protagoras of Abdera, known as the father of debate, has gained recognition for introducing debate in education over 2,400 years ago in Athens, Greece (Darby, 2007). Snider and Schnurer (2006) reported that Confucius and other Chinese philosophers wrote famous treatises that described debate as an effective method of learning. Darby (2007) pointed out that debate as a teaching method flourished throughout the nineteenth and early twentieth century and then waned in popularity. In the 1980s, stimulated by the philosophy of promoting critical thinking, interest in debate as an educational teaching strategy was rekindled again (Darby, 2007).

Debate activities are commonly conceptualized in relation to competitive debate, often in the form of well-established debate leagues and competitions (Jones, 1994). The majority of the studies that have investigated debates so far seem to revolve around L1 competitive debates whose benefits are well documented (e.g., Mezuk et al., 2011). Bellon (2000) wondered to what extent the existing research on competitive debate supports the implementation of debate in class. In other words, Bellon questioned the possibility of achieving the benefits gained from competitive debate with in-class debates. Bellon noted that in-class debates stand a big chance of reaping these benefits. To support this view, Bellon drew on cognitive research and constructivist literature that outlined a number of pedagogical characteristics that are conducive to effective learning, such as the provision of ample opportunities for oral communication. Bellon argued that debates that are designed "in accordance with general principles of good teaching" (p. 169) can embody these characteristics. Kennedy (2007) and Worthen and Pack (1992) maintained that debates can be easily integrated into the school curriculum. Tumposky (2004) pointed out that there is already a "widespread acceptance of debate as a vehicle for teaching and learning at the levels of both secondary and higher education" (p. 52).

One area of education that can profit from adopting debate as a teaching tool is language teaching. There is a great deal of consensus in the literature that in-class debate activities hold promise as a conducive mechanism for language learning (e.g., Alasmari & Ahmed, 2013;

Darby, 2007; Lieb, 2007; Rybold, 2006; Snider & Schnurer, 2006; Zare & Othman, 2013). Stewart (2003) stated in this regard that by "using debate, ESOL teachers can direct learners' study and practice of listening, speaking, reading, and writing skills, in combination with critical thinking skills" (p. 10).

2.3 Benefits of in-class debates

Debate seems to provide fertile ground for reaping many benefits. Literature is replete with benefits that are associated with debate participation. I confine myself to discussing the benefits of in-class debates (and not competitive debates) reported in L1 and L2 debate literature. I will focus on the benefits that are widely reported, so the review and the benefits presented are by no means exhaustive nor comprehensive since this is beyond the scope of this chapter, and it is also extremely difficult to list all the benefits. However, two major well-documented benefits—improvement of language and argumentation skills—will receive extensive attention since they form the basis for this PhD study.

2.3.1 Public speaking

Many people are terrified of public speaking. Rybold (2006), for instance, mentioned that Americans rank the fear of public speaking even above death. Fear of public speaking can be a significant barrier to effective communication. Freeley and Steinberg (2005) regarded debate as an ideal tool for aiding students to develop coping mechanisms that help them to manage anxiety during speech delivery. They further argued that through debate preparation, students develop confidence about the content they want to convey. So, through preparation and participation in multiple debates, students develop confidence and learn to cope with the inevitable nervousness caused by debate. This benefit is reported in many studies. For example, in Zare and Othman's (2015) study, the respondents reported that debates helped them to reduce their stage fright. They stated that they were initially anxious and afraid of talking in front of other students and that these feelings were weakened after participating in debates. Accordingly, they gained self-confidence and managed to overcome this fear as this respondent recounted: "It reduces the level of my anxiety in speaking in front of the crowd. I guess, in order for one to have the confidence in speaking, they need to experience it" (p. 164). Oros (2007) also remarked that debates helped many students to overcome the fear of public speaking.

2.3.2 Disciplinary learning

Debate-like activities can also enhance disciplinary learning. Goodwin (2003) argued that "teaching experience as well as empirical research affirms that debating helps students develop

content mastery" (p. 157). Students reported in his study that debates enabled them to think all week long about the material, delve deeper into the issues under consideration, develop a detailed and deep view of their readings, and consider a wide range of viewpoints. Similarly, in Kennedy's (2009) study, the participants reported greater gains in knowledge about the issues on which the debates centered. Scott's (2008) participants also revealed that debates helped them to acquire new knowledge, to gain an understanding of the topic, and to gain additional knowledge on the subject. Bellon (2000) pointed out that if students are not given a chance to debate about important concepts they are presented with in class, they will not be able to "develop deep or mature understandings of course content" (p. 172).

2.3.3 Citizenship skills

Oros (2007) and Zorwick and Wade (2016) contended that in-class debates can develop citizenship skills that are essential to achieve a resilient democracy and civil society. They stated that encouraging classroom participation in which pertinent political and societal issues are debated is likely to culminate in citizen participation in the long run. Rutten and Soetaert (2013) noted that debating and arguing competencies lie at the heart of a healthy democracy. They stated that having citizens who are "able to argue and take part in public debate seems to be an essential ingredient of a healthy democracy and an a priori powerful tool to become a democratic 'citizen'" (p. 3). Zorwick and Wade (2016) maintained that debate promotes democratic citizenship through contributing to the development of critical thinking and communication skills. Further, debate instils a sense of inclusiveness into students through giving them "'voice'—a way to interact with and to feel a part of the bigger world around them" (Martens, 2007, p. 4).

2.3.4 Other benefits

It was also reported in the literature that the use of debate as a teaching tool can yield many other benefits. One of these benefits concerns tolerating diversity in thinking. Many studies revealed that debate trains students to cope with opposing viewpoints and perspectives characterizing our daily life. It nurtures in them the predisposition to tolerate opposing points of view and consider other alternatives (e.g., Jackson, 1973; Kennedy, 2009). Another benefit concerns evaluative skills. Omelicheva and Avdeyeva (2008) mentioned that debates foster evaluative skills, as debates encourage students to weigh up facts, compare arguments, and expose fallacies in their opponents' line of reasoning as well as their own positions. In addition, Kennedy (2007, 2009) and Garrett et al. (1996) pointed out that debate develops empathy because through investigating the two sides of a controversial topic, students become more open-minded and better able to understand another person's viewpoint. Furthermore, Brown and Wilson (2016)

stated that debate stimulates students to reflect on their own performance and gain confidence in citing relevant resources. The other benefits that are widely reported include communication skills (e.g., Bellon, 2000; Zare & Othman, 2013), particularly oral communication skills (e.g., Budesheim & Lundquist, 1999; Kennedy, 2007), social and business skills (e.g., Bellon, 2000), and thinking on one's own feet (e.g., Randolph, 2007).

2.4 Argumentation and critical thinking skills

We live in a world where news and information travel at an astonishing pace. Technological development and social media in particular have added new channels that have accelerated the dissemination of this information at an unprecedented speed. In the face of these challenging circumstances, inspecting information with a critical mind and nurturing the ability to voice a balanced and well-grounded stance have become more important than ever. Educational methods must therefore cope with these challenges (Snider & Schnurer, 2006). Davis et al. (2016) also pointed out that the twenty-first century requires a reorganization of educational systems to offer students the skills needed to survive and thrive in the new diverse, global, and digital world.

The ability to think critically about everyday issues is arguably one of the most fundamental skills underlying success in everyday life as well as academic and professional contexts (Baron & Brown, 1991; Butt, 2010). Critical thinking helps in keeping an open mind and making better decisions. Good thinking skills also enable people to critically verify incoming information, consider alternative viewpoints, and produce counterarguments (Osana & Seymour, 2004). In addition, critical thinking stimulates going beyond challenging an opposing view in an attempt to develop a new perspective and make speculations about the future (Roy & Macchiette, 2005).

Osana and Seymour (2004) maintained that given the intricate nature of problems in virtually all everyday situations, nurturing students' ability to make critical judgments and evaluations is a crucial responsibility that needs to be assumed by the educational system. So modern education should take the responsibility to help students "to use their minds well, in school and beyond" (Kuhn, 2007, p. 110). Therefore, searching for pedagogical tools that infuse students with the ability to reason with critical thinking and argue with credible evidence is an overarching priority.

Many studies have voiced concern about students' ability to think critically (Scott, 2008; Willingham, 2007). Many of these studies have linked deficiencies in critical thinking to a lack of instruction in critical thinking skills rather than to an inherent inability of students to acquire such skills (Butt, 2010; Worthen & Pack, 1992). To improve students' critical thinking, Gadzella et al. (1989) proposed teaching them the skills of reasoning and then involving them in activities

that specifically require these skills. Yet, instructors face the challenge of finding appropriate resources and employing the appropriate pedagogy to use and develop critical thinking skills (Camp & Schnader, 2010; Goodwin, 2003; Scott, 2008). Kuhn (2001) and Kuhn and Udell (2003) posited that exercises involving thinking and reasoning skills in a cognitively rich context suffice to progress these skills.

Students' deficiency in critical thinking is particularly noticeable in their limited ability to support reasons with evidence (Yeh, 1998). Empirical evidence has revealed that students struggle with providing adequate justifications for their claims (Bell, 2004; Hsu et al., 2015; Walton, 1996) and generating counterarguments to rebut the opposing side (Sadler, 2004; Stapleton & Wu, 2015). Weaknesses in argumentative reasoning can have serious implications and therefore warrant nurturing (Crowell & Kuhn, 2014; Kuhn, 1991; Kuhn & Udell, 2003; Yeh, 1998).

The ability to generate and evaluate arguments has been widely deemed to be essential to good thinking (Mercier, 2011; Mercier & Sperber, 2011) and of critical importance in the twenty-first century (Crowell & Kuhn, 2014). Argumentation skills are important for both in and out of school. In school subjects, students are often confronted with issues and positions that need to be argued, defended, or evaluated (Means & Voss, 1996). Out of school, students encounter controversial issues in the media and on the Internet. Therefore, it is important more than ever that students make reasoned and informed decisions. To do this, they should be able to evaluate the strengths and weaknesses of their own and others' standpoints and to support their positions with relevant and adequate evidence (Marttunen & Laurinen, 2001). Yet, little is known in the educational literature about how to foster students' argumentation skills (Crowell & Kuhn, 2014).

2.4.1 In-class debate and argumentation skills

Given the importance of fostering students' argumentation skills, it is imperative to devise activities and create an environment facilitative of meeting this goal. The debate process is believed to offer one of the best methods for training and developing reasoning/argumentative skills (Butt, 2010; Oros, 2007; Roy & Macchiette, 2005; Zorwick & Wade, 2016). Debate, by its nature, is filled with opposing viewpoints. It puts debaters face to face with plenty of conflicting facts, assumptions, and perspectives that entail higher-level reasoning strategies. Involvement in debate, therefore, forces students to critically analyze the opposing side's reasoning and evidence, endeavoring to point out weaknesses and inadequacies in them. Debate pushes debaters to search, identify, inspect and evaluate arguments, overcome personal prejudices and biases, and identify inconsistencies and inadequacies in opponents' line of reasoning.

2.4.2 Research on debate-argumentation effects

Some studies attempted to capture the effects of debating on argumentation skills. They reported improvement in the argumentative skills of students after taking part in debates. However, the majority of these studies were based on students' self-reports and instructors' observations and were conducted in the L1 context. For example, Zorwick and Wade (2016) reported that there was unanimous agreement among social studies/history teachers (participants) that debate activities enhanced their students' ability to "write arguments to support claims using valid reasoning and relevant and sufficient evidence" (p. 441). Ninety-eight percent of these teachers felt that the participating students improved in their ability to "gather relevant information from multiple authoritative print and digital sources" and "draw evidence from informational texts" compared to the students who did not participate. Kennedy (2009) pointed out that debates helped her students to become familiar with the arguments of the two sides of the debated issues, and because of this, 57.5% of the participants had changed their views by the conclusion of the debates. Oros (2007) reported that the students that participated in the Introduction to World Politics course (with debates) produced far more arguments and with a higher degree of support in the final essays than the students who participated in the Introduction to Political Science course (with no debates). Oros attributed this difference in part to the debate experience. He noted that this difference persisted after the conclusion of the courses.

The studies that have investigated the argumentation-debate relationship in the L2 context are very scarce. The few existing studies have massively recognized that debate promotes argumentation skills (e.g., Krieger, 2005; Lieb, 2007; Lustigova, 2011; Zare & Othman, 2013, 2015). However, much of the advanced evidence correlating debate with argumentation enhancement in this context is anecdotal, based on participants' reports and instructors' observations. For example, Zare and Othman's (2015) participants reported that debates helped them to use evidence and data to support their arguments and viewpoints with a mean of 4.13/5. Also, Davidson (1995, as cited in Krieger, 2005) found that "many [Japanese] students show obvious progress in their ability to express and defend ideas in debate . . . and [recognize] the flaws in each other's' arguments."

On the whole, though there is wide recognition that debates promote argumentation skills, we still lack an empirical verification of this recognition. Most of the research that we have so far is anecdotal, is based on students' perceptions and instructors' observations, and is conducted in the L1 and higher education contexts. Therefore, empirical research is needed to establish a direct causal link between debates and argumentation skills (Camp & Schnader, 2010;

Randolph, 2007). Such an empirical evaluation would provide teachers, researchers, and policymakers with important guidance regarding their efforts to foster students' argumentation competence (Yeh, 1998).

2.5 In-class debates and language proficiency

This section reports the results of previous research conducted to study the effects of debates on language proficiency. As is the case with the debate-argumentation relationship, the vast majority of past research in this area was conducted in the L1 context. Research around the effects of debate in the L2 context on language proficiency is in its infancy. Brown (2009) remarked that unlike in the L1 context, debates as a teaching tool are generally absent from L2 curricula. Nevertheless, Brown noted that debates, as in the L1 context, can help L2 learners improve their command of language.

Akerman and Neale (2011) argued that there is a strong body of evidence that correlates L1 debate participation with improvements in academic achievement. They maintained that different studies have revealed that debate has a substantial bearing on the attainment of young people irrespective of their backgrounds, especially on the improvement of their literacy skills.

Mezuk et al. (2011) conducted a large-scale research—the largest evaluation study of an L1 debate program on achievement ($N = 9145$, over a 10-year period)—exploring the relationship between participating in debates and high school graduation rates in a public school in Chicago. They found that 90% of debating students graduated from high school over the study period, as compared to 75% of comparable nondebating students. Also, the debating students were 19% more likely to graduate from high school than the nondebating students. Futhermore, their study revealed that the debating students significantly outperformed the nondebating students in reading and English tests, gaining on average an additional 1.02 and 1.04 points, respectively. These findings, according to Mezuk et al., "suggest that debate programs may offer a means to extend learning time and promote engagement with scholastic materials in a manner that translates into academic performance" (p. 362). Anderson and Mezuk (2012) also reported that the debating students were 3.1 times more likely to graduate from high school than the nondebating students and that they scored substantially better in reading and English tests.

In a study undertaken by Littlefield (2001), high school students in America were asked to identify the perceived benefits gained from participating in debates. Communication/speaking skills were ranked top, accounting for 22% of the benefits. The participants also reported that participating in debates enabled them to have "more knowledge than their peers and [be] better at reading, writing, and speaking" (Littlefield, 2001, p. 90). In the same vein, Collier's (2004) study revealed an improvement of 25% in the reading scores of 209 high school debaters in

comparison to 212 randomly selected nondebaters from the same schools (see also Breger, 2000; Williams et al., 2001; Zorwick & Wade, 2016).

In the L2 context, debate research is characterized by the predominance of anecdotal evidence. Lustigova (2011) conducted a study in which 475 Czech students from intermediate and lower-level classes were engaged for one semester in simplified debates. The English course, in which these debates were incorporated, was a required course for a B.A. degree. Lustigova pointed out that the participants considerably improved in terms of the amount of time for which they were able to speak. At the start of the debate sessions, the debates lasted approximately six minutes, while at the end of the semester, they exceeded 20 minutes. Moreover, Lustigova observed that the students' four language skills, in addition to their vocabulary, had substantially improved by the end of the semester:

> Taking into account the four language skills, the students manifestly improved their writing, reading, and listening skills, as was evidenced within the classroom with enhanced performance in written homework and oral consultation sessions. Vocabulary tests, and other oral activities, during the semester also demonstrated a sufficient increase of gained knowledge. (p. 25)

Also, in Aclan and Aziz's (2015a) study, the participants reported improvement in the four skills after participating in debates. It is noteworthy that the majority of the studies that investigated the effects of L2 debates on language proficiency focused on speaking.

As to vocabulary, Aclan and Aziz (2015b) conducted a qualitative study to explore why and how FL students learn vocabulary in classroom debates. The participants reported improvement in their vocabulary; they related this improvement to "debate's interactive nature requiring contextualized and meaningful language use from preparation to actual debate" (p. 102).

To sum up, so far research demonstrating how participation in debate results in language development has been rarely undertaken (Omelicheva & Avdeyeva, 2008). Littlefield (2001) noted that this dearth of research is particularly noticeable in the secondary school context. It manifests itself in the fact that "very few manuscripts dealing with high school debate have been published in academic journals" (Littlefield, 2001, p. 83). The scarcity of research that Omelicheva and Avdeyeva, and Littlefield pointed out concerns the L1 context. In the L2 context, the debate research is scarce in the extreme (e.g., O'Mahoney, 2015). Inoue and Nakano (2004) stressed that the existing anecdotal evidence, which is mainly based on surveys, "needs to be tested for validity" (p. 8). In other words, without empirical evidence the claimed effects of debate remain groundless.

2.6 Criticisms against debates

Many scholars have leveled a number of criticisms against debates. Critics have mainly charged that debates promote dualism and negative emotional qualities like hostility and fighting, put too much focus on competition, and do not lead to rich negotiations. In this section, I discuss these criticisms and respond to them.

Tumposky (2004) stated that debates reinforce a bias toward dualism and marginalize the multiplicity of perspectives. Tumposky further maintained that the dualistic format inherent in debates often leads to trivializing complex ideas and focusing on aspects that strengthen one's side. Similarly, Tannen (1998) pointed out that arguing further reinforces debaters' contention through focusing on each other's weaknesses in logic and trying to twist facts to tip the scale in their advantage rather than fairly endeavoring to settle the differences.

However, a number of researchers do not support this view. Jackson (1973) stated that "inherent in debate is the opportunity for the student to be placed in a situation which calls for tolerance of opposing points of view" (p. 152). Also, MacArthur et al. (2002) contended that debate holds promise as a means of enabling students to learn about multiple perspectives on controversial issues. Scannapieco (1997) mentioned that 76% of dentistry students surveyed "agreed that participation in the debate helped them to realize that most issues are not clear cut" (p. 960). Kennedy (2007) found that debating helped her students develop empathy as this student said:

> When you went to the debate you listened to both sides of the argument, which I thought was the main strength of the debates, that you do see both sides, rather than just seeing it from one point of view. (p. 184)

In another study, Kennedy (2009) reported that between 31% and 58% of her participants changed their views after taking part in, or observing, each debate. In Fallahi and Haney (2007), 80% of debaters reported that debates changed their personal viewpoint on the topic under debate. Similar results emerged from Simonneaux (2001), whose students in his biotechnology classes reported a change in their opinions after participating in debates. To mitigate bias toward dualism, Budesheim and Lundquist (1999) proposed requiring students to defend a side they personally oppose.

Tannen (1998) and Trapp (1986) stated that students may link debate to negative emotional qualities like hostility and fighting since debating involves open disagreement. However, in Goodwin's (2003) study, only four students (out of 52) expressed concern about hostility, fighting, or other negative emotional consequences of debate activities. Bellon (2000) said that debates equip debaters with the essential argumentative skills needed to resolve (intellectual)

conflicts without making recourse to verbal aggression. He further noted that debates enable debaters to understand others' line of reasoning even if they strongly disagree with it. In Fallahi and Haney (2007), 74% of the participants believed that they would be more likely to experience negative emotions, such as anger, frustration, and nervousness "if they were to share their personal feeling in the classroom during group discussions or lectures, rather than debate an assigned standpoint" (p. 86).

Tannen (1998) and Tumposky (2004) criticized debates for the emphasis they give to competition. They argued that debates create a confrontational environment, which sometimes makes debaters feel uneasy. This applies, according to them, in particular to females who would rather seek compromise and consensus building than public argument. They also added that students tend to trivialize issues for the sake of winning. However, many studies have attested that the competitive nature of debates does not downgrade the educational benefits (e.g., Butt, 2010; Colbert, 1987; Jones, 1994; Wood & Rowland-Morin, 1989). Wood and Rowland-Morin (1989) stated that "a student involved in debate may find the competition rewarding (extrinsic interest) but that does not diminish de facto the educational motivation (extrinsic motivation)" (p. 82). Goodwin (2003) elicited evaluations about debate-like activities through essays in which participants were asked to articulate their views regarding whether the competitive nature of the debates intimidated or silenced them. The thematic analysis of the responses revealed that only two students (out of 52) voiced discomfort with the competitiveness of debate. A similar finding also emerged from the studies of Brown (2015) and Hall (2011).

Crowell and Kuhn (2014) maintained that debaters may initially focus on winning a debate, but with time this focus shifts towards the quality of argumentation itself, including generating solid counterarguments and rebuttals. Importantly, instructors need to underscore that debate is a competition of persuasion and is not about winning. It is true that some students may be preoccupied by the thought of winning the debate, or paralyzed by the idea of losing it, but this can be diluted by emphasizing that debate is meant to enrich their learning experience and not to yield a winner. Merrell et al. (2017) recommended in this regard against awarding

> additional points for 'winning' the debate … removing an incentive for winning should remind students that their goal in the debate is not to beat an opponent, but rather to demonstrate that they can construct, apply, and defend against well-supported arguments. (p. 65)

Lastly, Pica (1996) maintained that reflective tasks like debates do not lead to much negotiation because mostly assertive learners dominate the interaction at the expense of less assertive learners. However, well-structured debates that are staged in accordance with formats that level

the playing field in terms of the amount of time allocated to each participant can yield fruitful negotiations. Additionally, because of their reflective feature, debates can push negotiation to new heights, since debaters tend to challenge each other's arguments and the way they frame them, especially during rebuttals. On top of that, research has shown that debates encourage even the passive and less assertive students to take part in interaction (Dundes, 2001; Stewart & Pleisch, 1998).

It should be noted that I am not denying the existence of the listed concerns, but I take the view that there is no conclusive evidence indicating that they exist on such a scale that they jeopardize the effectiveness of in-class debates. For example, Goodwin reported that 25 students (out of 52) pointed out a number of negative features of debates. However, these features were diverse and none of them was raised by more than four students.

Nevertheless, I should admit that in-class debate implementation poses some challenges. One of these challenges concerns class size. Many studies (e.g., Hyde & Ruth, 2002; Musselman, 2004) reported that it is quite difficult to avail all students of the opportunity to actively participate in each debate, especially in large classes. To tackle this pitfall, the debate literature proposed some debate formats that involve all students in a certain way in the debate. For example, Musselman (2004) required her students not directly participating in the debate to write debate points, a piece of writing in which they argued against or in favor of one side of the debate. Moreover, many debate scholars (e.g., Goodwin, 2003; Musselman, 2004) remarked that their students expressed discomfort with the labor-intensive preparation required by debates. Preparing for debates entails reading sources, writing essays, scheduling meetings with instructors, etc. This can be daunting and may lead to frustration (Hall, 2011). To lessen this frustration, Hall (2011) suggested offering students enough preparation time and scheduling debates in "non-heavy test/assignment times."

A final challenge worth mentioning concerns the massive time investment instructors have to make to develop, maintain, and grade debates. For instance, Musselman (2004) mentioned that she usually needed about four hours per debate to read primary sources, send e-mail feedback, and arrange meetings with students. These concerns were also echoed by Jackson (1973), yet he downplayed them given the plenty of benefits that can be obtained from debates. Similarly, Darby (2007) maintained that "although there are disadvantages to using the debate as a teaching-learning strategy, the benefits far outweigh the disadvantages" (p. 9). Therefore, Jackson and Darby noted that a teacher should not eschew debates simply because of the extra effort they require, because "good teaching always requires extra effort" (Jackson, 1973, p. 154). In a similar vein, Oros (2007) argued that though in-class debates entail extra effort, "the potential payoff is great" (p. 309).

2.7 In-class debates in teaching practice

In general, the bulk of the available instructional materials and procedures regarding in-class debates are developed for the L1 context. The L2 debate literature is scarce. Morse's (2011) account is illustrative of this scarcity: "The materials that describe debate as a part of the second or foreign language learning curriculum are extremely limited" (p. 108). The few existing works in this context provide no or few details about how debates are staged in practice. In this section, I discuss how different elements of in-class debates have been implemented in teaching practice. The discussion revolves around how debate topics are selected, debate preparation, how students are assigned to teams and debate sides, whether they are tasked to research one or both sides of the debate, debate assessment, and debate formats. I discuss the implementation of these elements in relation to the L1 and L2 contexts only in case there are relevant differences. This section also seeks to identify design elements that can inform the construction of a workable debate task design (in the next chapter).

2.7.1 Debate topics

There seems to be a consensus among debate scholars that one of the critical conditions that need to be met for debates to be effective concerns selecting appropriate topics. Research has demonstrated that topic interest is an affective variable that enhances learners' involvement in the learning process and positively impacts their language development (Lee & Pulido, 2017). An interesting debate topic can intrigue students and engage them in the debate process (Snider & Schnurer, 2006). Conversely, an uninteresting topic can lead to frustrations and discourage students from getting involved in the learning process (Zare & Othman, 2015). Therefore, it is important to facilitate the selection of topics that students consider relevant and can relate to, as this can even shape students' attitude towards the whole debate process (see Dörnyei, 2002; Dörnyei & Kormos, 2000). Wade (1994) reported that 93% of her participants expressed a willingness to participate in debate, provided that they are presented with interesting discussion topics. In most of the debate literature (L1 and L2) reviewed, debate topics were selected by instructors. Their selection was mostly driven by course content. Nevertheless, Zare and Othman (2015) recommended involving learners in the process of selecting debate topics to enhance participation during debates.

2.7.2 Debate preparation

Debate literature provides strong evidence that underscores the key role sufficient research plays in boosting students' self-confidence and performance during actual debates (e.g., Ericson et al., 2003; Snider & Schnurer, 2006). Research enables students to gather solid arguments,

inspect 'the other' perspective, and find appropriate words to express their stance persuasively. Just floating around fuzzy ideas does not lead to a proper debate. Research can be conducted in the form of reading books, articles, research reports, magazines, newspapers, etc. Students need to assemble evidence in the form of statistics, experimental data, common knowledge, and pertinent examples (Lubetsky et al., 2000). Preparation requires in the first place a thorough understanding of the issue at hand, since, as Hensley and Carlin (1994, as cited in Morse, 2011) emphasized, "a team can't debate, or even research, without a basic understanding of the topic and its issues" (p. 28).

Darby (2007) mentioned that debate preparation also promotes student collaboration since "students must then engage in constructive teamwork to unify their position and eliminate redundancy" (p. 1). Importantly, the participants in Aclan and Aziz's (2015b) study identified preparation for debate as an important factor that helped them to improve their vocabulary.

2.7.3 Assigning students to teams and debate sides

The L1 debate literature shows many variations in how students are assigned to teams. In Brown (2015), students were allowed to choose the side they wanted to defend. Similarly, Omelicheva (2006) formed debate teams taking into account students' preferences for the different positions in the debate. Omelicheva stated that allowing students to select a negative/affirmative position in a debate increases their interest and satisfaction with the debate experience. In contrast, Budesheim and Lundquist (1999) required their students to defend a position inconsistent with their opinion. They did this to reduce students' tendency to blindly stick to their existing opinions and hence become tolerant to the opposing position. Musselman (2004) scheduled six debates in her course. Each student in the course took part in two debates as an antagonist (defending the affirmative or negative position), two as a questioner (challenging the antagonists and conciliators) or conciliator (offering a compromise or alternative position), and in two debates as an author of a debating point (writing one or two paragraphs in which they defended a side in line with their opinion). In Walker and Warhurst (2000), two teams (for and against) of between four and six students argued against each other, trying to convince a third group (audience group) of the legitimacy of their position. The "audience" group of students were required to adjudicate after considering the persuasiveness of the advanced arguments, team effort, and the quality of presentation skills.

Unlike the L1 debate literature, the L2 literature provided few details about how debates are staged in practice. For example, Aclan and Aziz (2015b) did not reveal anything about how debates were carried out in their study. Lustigova (2011) stated that "group members were assigned by the teachers or the students determined the groups by themselves" (p. 21), but she

did not explain how this happened and whether the students were allowed to choose sides. Rashtchi and Sadraeimanesh (2011) claimed that they used debate as a teaching tool in their study, but it appears from their description that they used classroom discussions.

Students were not always happy with being forced to defend a side of a debate inconsistent with their standpoint. Fallahi and Haney (2007) reported that half of their students "felt frustrated at some point during the debate because they were not able to present points that were in agreement with their beliefs" (p. 86). Randolph (2007) pointed out that 40% of her students did not like being randomly assigned to groups and noted that they would have staged better debates had they been allowed to choose their teams.

2.7.4 Researching sides

Students were often asked to prepare both sides of the debate; they were assigned to the pro or con side on the day of the debate. In asking students to research both sides of the debate, instructors would like to offer their students the opportunity to strengthen their own position by having evidence and an understanding of the opposing position (Budesheim & Lundquist, 1999; Oros, 2007; Zare & Othman, 2015). In some other studies (e.g., Brown, 2015; Goodwin, 2003), students were required to research one side of the debate. Students were then instructed to defend their side by finding the arguments and evidence that support it.

2.7.5 Debate assessment

Brown (2015) and Kennedy (2007) advocated utilizing student assessment of debates in order to increase student involvement. Merrell et al. (2017) recommended using grading rubrics that include both preparation and in-class performance. Smith (1990) found that student and instructor ratings correlated highly and that students experienced evaluating other students as a valuable experience. Garrett et al. (1996) proposed grading on a pass/fail basis to reduce the anxiety associated with debate as an unfamiliar activity and to encourage students to see it as a cooperative endeavour. Walker and Warhurst (2000) stated that students can work together to set up the assessment criteria for debates. However, they reported that their students were not comfortable with summative peer-assessment, and therefore they proposed using formative peer assessment. In Fallahi and Haney (2007), students received two grades for each debate: an individual grade assigned by the instructor based on performance during debate and a written summary of the team's research, and a team grade provided by every member of the team. In Hall (2011), students' performance in debates was assessed by faculty members. The assessment of debates was mostly based on the quantification of the scores of different components of debate (e.g., speech, arguments, evidence, and rebuttals) with a rating scale (see, for example, Hall, 2011; Omelicheva, 2006).

2.7.6 Debate formats

A number of L1 in-class debate formats have been proposed in the literature. Most of these debate formats are closely modeled on, or are identical to, competitive debate formats (Merrell et al., 2015). Moreover, in most of these debate formats not all students are actively involved in each debate (Kennedy, 2007). In most cases, nondebating students fulfilled other roles in the debate. For example, in Musselman's (2004) study, the nondebating students took the roles of questioners and conciliators. The questioners challenged an antagonist with prepared questions; the conciliators thought up a compromise or an alternative solution, and the remaining students wrote an argumentative one or two paragraphs defending one side of the debate.

Here I present some examples of the debate formats used in the L1 debate literature. Kennedy (2007) proposed four debate formats (some with two versions) that can be used for in-class debates: fishbowl debates, think-pair-share debates, Lincoln-Douglas debates, and problem-solving debates. For example, in one version of a fishbowl debate, the teacher splits the class into two groups, and each group generates arguments for their assigned position. After each side has presented their arguments, the groups present their rebuttals. Darby (2007) employed a debate format comprising three phases. In the first phase—*constructive argument*—each side presents their constructive arguments without interruption. In the *rebuttal* phase, each side is given the opportunity to challenge and rebut the constructive arguments. In the *class interaction* phase, each nondebating student receives the opportunity to get involved in the debate by asking questions they had to prepare, making observations, signaling areas of potential compromise, or pointing out alternative positions that were not discussed during the debate. Oros (2007) presented a debate format that accommodates two two- or three-member teams and allows each debate to last 18 minutes—three three-minute rounds for a "pro" and a "con" team. In the first two rounds, arguments supporting each side are presented, while in the last round each team rebuts its opponents' arguments. After each debate, a minimum of 15 minutes of class discussion takes place to engage other students beyond the four to six students that participated in the debate.

Because almost all the available debate materials were designed for the L1 context (Lieb, 2007; Stewart & Pleisch, 1998; Zare & Othman, 2013), it is not surprising that L2 instructors mostly drew on these sources when selecting debate formats. For example, Zare and Othman (2015) used a classroom debate format modelled on the British Parliamentary debate, in which students are randomly assigned to two teams (government and/or opposition). The first speaker for the government opens the debate by defining the debate topic (the motion) and subsequently presenting his/her speech. Next, the first speaker for the opposition side rebuts the arguments of the government's first speaker. After that, the second speaker of the government

team rebuts the arguments of the opposition's first speaker and presents his/her arguments and speech. The debate continues in this fashion until all speakers from both sides have presented their cases. The final speakers from both sides are allowed to make concluding statements and end the debate. During the debate, all speakers and participants are allowed to offer Points of Information (POIs) to their opponents.

On the whole, the vast majority of debate procedures that provide guidelines about how debates can be staged in teaching practice come from the L1 literature. Procedures geared to the L2 context are very scare. Moreover, all of the few studies that described debate implementation in the L2 context were conducted in higher education. This hinders an effective implementation of debates in the L2 secondary school context, which is the primary focus of the current research. In short, the challenges facing the implementation of in-class debates in the L2 secondary school context can be summarized as follows:

- The absence of a clear debate task design that informs students how to make the most out of debates.
- Not every student is equally involved in debating. Mostly self-assertive and language-proficient students dominate debates.
- The absence of a clear-cut debate structure (formats) suitable for the L2 secondary school context.
- Students are not guided to make proper preparation. This leads to superficial debates in which weak and spurious arguments based on feelings are used.

Given that the main purpose of this study is investigating the effects of L2 in-class debates on secondary students' language proficiency and argumentation skills, developing a comprehensive debate task design that addresses these challenges is of paramount importance. Without this design, the potential learning outcomes of debates are unlikely to be realized (Omelicheva, 2006; Zare & Othman, 2015). The next chapter will be dedicated to fulfilling this aim.

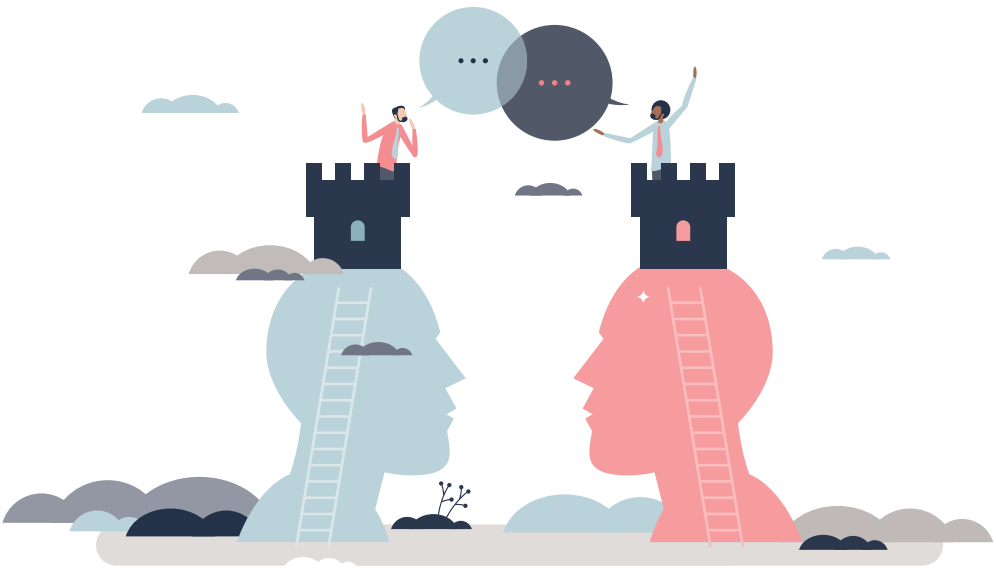
2.8 Conclusion

In this chapter, I have mainly discussed the potential effectiveness of in-class debates and a number of benefits they can yield. I have concentrated on two benefits—improvement of language proficiency and argumentation skills—because they form the basis for this PhD study (see Chapter 3.3.1). It is important to reiterate that the published debate literature has primarily used student perception data to establish these benefits and other benefits of debate (Dy-Boarman et al., 2018). To objectively evaluate the claimed and perceived effects of in-class debates, it is essential to conduct debate interventions.

As we have seen, most of the available debate materials and procedures have been developed for the L1 context (e.g., Lieb, 2007; Zare & Othman, 2013), which drew heavily on competitive debate (Merrell et al., 2015). However, to investigate the potential causal link between debate and improvement of language and argumentation skills in the L2 secondary school context, it is imperative to engineer a comprehensive debate task design that is grounded on meaningful exchange of arguments, that can ensure optimal preparation and participation of all students (Brown, 2015), and that takes into account the particularities of the Dutch secondary school context. Bellon (2000) stressed that such a design must be premised on pedagogically sound ground to leverage the potential of debates. The next chapter will be dedicated to the construction of this task design.

Chapter 3

Design and Evaluation of a Debate Task Design



"Incorporating debate in the foreign language classroom can take on many forms, and one is not necessarily better than another. Rather than adopting a one-size-fits-all method, program dynamics … should factor into choosing a suitable approach." (Brown & Bown, 2014, p. 12)

Abstract

This chapter discusses the debate task design that was built in accordance with the principles of educational design research (EDR). This study proceeded through three main stages. In the preliminary analysis stage, a context analysis with the aim of scoping out the educational problems as well as the pedagogical potentials of debate (that respond to these problems) was conducted. Then, a literature review was undertaken to inform the construction of the conceptual framework of the study on the basis of which the initial prototype of the study was developed, and through iterative cycles of testing and refinement the final prototype was finalized. The final prototype was piloted.

The chapter is structured in the following way. After the introduction, I define EDR and discuss its main characteristics and stages. Then, I address context analysis. Next, I present a literature review that covers two main areas of research (task-based language instruction and debate literature) leading to the identification of a number of empirically underpinned design principles that guided the construction of the initial prototype. This prototype is delineated in the subsequent section. Then, the prototyping phase, consisting of cycles of design analysis and redesign, is covered in Section 4. The final part of this chapter discusses the role of the instructor in the intervention and offers a reflection on the whole intervention.

3.1 Introduction

In Chapters 1 and 2, I argued that in-class debates may hold the potential to serve as a conducive mechanism for second language (L2) learning. Despite their potential benefits, in-class debates have been rarely employed in L2 language classrooms (e.g., Brown, 2009; Shamsudin et al., 2017). Of those who used in-class debates, only few of them implemented debates in a structured way, and even fewer integrated them systematically into syllabi and curricula (Archer & Miller, 2011; Merrell et al., 2017; Oros, 2007). In short, existing research has touted debates as a potential effective L2 tool without elucidating how to implement them and empirically substantiating their effectiveness.

Many challenges hinder the implementation of debates in the L2 classroom, not least of which is the scarcity of materials specifically designed for L2 learners (e.g., Lieb, 2007; Zare & Othman, 2013). Probably the greatest challenge resides in the absence of balanced debate task designs that explicate the mechanics of the debate activity in a stepwise manner that leads to a rich, engaging, and dynamic learning experience (Brown, 2015). This dearth of research concerns in particular the L2 context (e.g., O'Mahoney, 2015; Zare & Othman, 2013, 2015), especially the secondary school context (Littlefield, 2001).

Brown and Wilson (2016) argued that debate design is an impactful variable that warrants careful attention. In a related vein, Budesheim and Lundquist (1999) maintained that an adequate design of in-class debates is essential to maximize their educational value. Bellon (2000) contended that such a design must be premised on pedagogically sound ground for debates to be effective. Without this design, the promised learning outcomes of debates are unlikely to be realized (Omelicheva, 2006; Zare & Othman, 2015). It is important to note that though there is some literature about first language (L1) competitive debate procedures, many of these are unsuitable for classroom environments (Merrell et al., 2015). Therefore, this chapter attempts to engineer an accessible and workable debate task design that lends itself to implementation in the L2 Dutch secondary school context.

Using the Educational Design Research (EDR) approach (McKenney & Reeves, 2012; Nieveen, 2007), we built a debate task design and refined it through a number of iterations. EDR provided the opportunity to approximate the ideal of the study through a number of iterative cycles of design, development, and prototype evaluation. Classroom observations, group discussions, questionnaires, and interviews guided the modifications made to the debate task design. These modifications, while guided by pedagogical goals, led to the construction of a debate task design that lived up to prespecified goals. The developed debate task design will provide us and instructors in other schools with the necessary instructional guidance that specifically explicates in a stepwise manner what needs to be done in every stage of the debate

task. A comprehensive and robust debate task design would eventually enable us to adequately and reliably answer the main research questions of the current dissertation. In short, the ultimate purpose of this study was to develop a debate intervention that allows us to gain insights into L2 secondary students' attitude towards in-class debate, elicit their perceptions regarding the effects of debate pedagogy on language skills, and gauge the extent to which this pedagogy can improve different dimensions of students' speaking, writing, and argumentation skills. The overriding research question posed in this study was:

RQ.1 What are the characteristics of an L2 debate task design that facilitates meaningful in-class debates and optimizes students' involvement in the learning process in the Dutch secondary school context?

3.2 Educational design research

Nieveen (2007) defined EDR as "the systematic study of analyzing, designing and evaluating educational interventions in order to solve complex educational problems for which no ready-made solutions are available and to gain insight into key design principles" (p. 89). Looking at EDR through a similar lens, McKenney and Reeves (2012) characterized it as an intervention and process-oriented approach to research that employs diverse methods to investigate the development and implementation of solutions to emerging educational problems. EDR falls within the broader category of design research. By its nature, design research fits the educational practice, as it attempts to develop research-based solutions and respond to practical needs in educational practice (Plomp, 2007). EDR is conducted to design and develop interventions when a need arises for which no practical and ready-made how-to-do guidelines are available (McKenney & Reeves, 2012; Nieveen, 2007; Plomp, 2007; Plomp & Nieveen, 2013). EDR is flexible (Reinking & Bradley, 2008) as well as iterative and cyclic in nature (Plomp, 2007); the process of analysis, design, evaluation, and redesign activities are repeated until the "realized outcomes" are close enough to the "intended outcomes."

EDR is grounded in theory and developed in cooperation with the targeted audience, who function as a source of input and feedback about the innovation (Kopcha et al., 2017). In this way, "design principles and learning theories are enacted and refined in local contexts" (Schmidt & Kopcha, 2016, p. 2336). EDR is responsive to students' learning styles (Kwon et al., 2013) and context particularities (Plomp, 2007; Plomp & Nieveen, 2013). Because of all these qualities, EDR is considered an appropriate research methodology for developing interventions serving pedagogical purposes (Ivey, 2013; Kwon et al., 2013; McKenney & Reeves, 2014). Along the same lines, Wallin and West (2013) argued that EDR "has the potential to bridge the gap between theory and practice and thereby improve teaching and learning" (p. 182). We believe

that EDR is congruent with the aims of this study and provides avenues for optimizing the implementation of debates.

Nieveen (1999) argued that high-quality interventions satisfy four generic criteria. First, the components of the intervention are built on state-of-the-art knowledge, i.e., the intervention is relevant (content validity). Second, there is a coherent and consistent link between all components, i.e., the intervention is consistent (construct validity). Third, end users (teachers and learners) consider the intervention to be practical and can conveniently use the materials as prescribed by the design researcher, i.e., the intervention is practical. Fourth, the desired outcomes have been realized to a satisfactory extent, i.e., the intervention is effective.

The research process in research design has been diversely conceptualized (e.g., Bannan-Ritland, 2003; McKenney, 2001; McKenney & Reeves, 2012; Reeves, 2006). Plomp (2007) argued that though different scholars have presented different conceptualizations of design research, they all agree that it goes through the following phases:

1. The preliminary phase, in which the context analysis, review of literature, and development of a conceptual or theoretical framework for the study are undertaken. This stage starts by identifying and formulating the problem of the study. Then, a review of literature is conducted with the aim of constructing a conceptual framework for the intervention. The findings of this stage guide the next stage of the design study (i.e., the prototyping phase).
2. The prototyping phase (the iterative design phase), which consists of iterations, each being a microcycle of research, with formative evaluation as the most important research activity aimed at improving and refining the intervention. This phase is guided by a preliminary prototype engineered in the preliminary phase. Each research cycle leads to a refined prototype based on results. This process is repeated until a final prototype embodying the intended outcomes is reached.
3. The assessment ((semi-) summative evaluation), which identifies the extent to which the intervention meets the predetermined goals.

Throughout all these stages, the design researcher needs to engage in systematic reflection and documentation (see Van den Akker, 1999). These are continuous reflective activities that need to occur during all cycles of the research. Cycles should be looked at as representing microcycles of research with a specific research/evaluation question (Plomp, 2007; Van den Akker & Nieveen, 2017). Plomp (2007) stated that there are many possible methods of formative evaluation at the researcher's disposal, including one-to-one evaluation, walking through with members of the target audience, field testing, trying out, etc. However, he maintained that it is

up to the design researcher to select the method that best fits the purpose of the research stage under study.

3.3 Preliminary analysis stage

The purpose of this phase is to identify and analyze the state of the problem with the aim of generating perspectives about how debates can be effectively implemented in the L2 classroom. This phase includes a context analysis and a literature review. Context analysis seeks to gain insights into the educational problem prompting this study and eventually formulate the intended outcomes that will steer the study. The literature review is aimed at selecting and reviewing relevant theories covering the assumptions necessary for a sketch to solve the problem at hand. The information derived from the context analysis and literature review will pave the way for formulating the design guidelines and specifications that will inform the construction of the initial prototype.

3.3.1 Context analysis

Context analysis is conducted to define the problem in its environment so as to shape an adequate understanding of it, identify potential causes, and gain insight into how to address it (McKenney & Reeves, 2012). John Dewey once said that "a problem well stated is a problem half solved." Therefore, it is imperative to purvey a thorough picture of the educational problem, explaining its root causes. McKenney and Reeves (2012) pointed out that there are three issues that need to be addressed at this stage: What is the current situation? What is the desired situation? And what is already known or suspected about the causes of this discrepancy? Context analysis, in brief, is meant to scope out the instructional problem and determine process and strategy to work towards the intended educational objectives. In this section, I shed light on the position of in-class debates in the Dutch secondary school context and discuss potentials and affordances of debate pedagogy that cater to key objectives pursued in L2 education, namely improving speaking, writing, and argumentation skills.

In-class debates in Dutch secondary schools

In Dutch secondary schools, debate is usually regarded as an extracurricular activity and not as a pedagogical tool that deserves a place in the regular curriculum (Gelinck, 2000; Van der Woude et al., 2011, 2012). Van der Woude et al. discussed a number of factors that obstruct the implementation of debate in Dutch secondary schools. First, debate is generally viewed as a game and not as a serious pedagogical tool. Second, there is an absence of a debate task design that informs teachers of the practical steps they should follow to engage students in debates.

Third, instructors are unaware of the potential benefits debate can confer on students. Van der Woude et al. called for research that demonstrates how debates can be adequately implemented in class and hence motivate instructors to employ them and ultimately experience their potential.

Research about debate as a teaching tool in the Dutch context is extremely scarce. Conducting scientific research that establishes how debates can be effectively staged and investigates their effects on language skills, particularly speaking and argumentation skills, would smooth the path for the introduction of debates in the Dutch curriculum (Bonset & Braaksma, 2008; Gelinck, 2000; Van der Woude et al., 2011, 2012). Therefore, van der Woude et al. called for more debate research in the Dutch context in order to convince teachers, policymakers, and stakeholders to welcome debates in Dutch secondary schools. They also contended that one way of investigating the effects of debate is through carrying out studies that compare debate-based instruction with traditional-based instruction. It is worthwhile noting that Van der Woude et al.'s and Gelinck's studies revolved around the use of debate in the Dutch L1 context; studies about debate in the Dutch L2 context are totally absent. To date, I am not aware of any study that has explored the implementation of in-class debates in the Dutch L2 context.

In-class debates and speaking skills

Speaking is mostly regarded as "the most complex and difficult skill to master" (Tarone, 2005, p. 485) and is generally considered the most important skill (Lustigova, 2011). Speaking a second or foreign language properly is a very intricate and arduous task considering the interwoven factors that come into play when acquiring this ability (Richards & Renandya, 2002; Romaña Correa, 2015; Shumin, 2002). Richards and Renandya stated that "a large percentage of the world's language learners study English in order to develop proficiency in speaking" (p. 201). Nonetheless, Baker and Westrup (2003) and Thornbury (2005) pointed out that speaking has been marginalized in the L2 class, while "it is a crucial part of second language learning and teaching" (Kayi, 2006, p. 1). As a result, it is no wonder that many studies (e.g., Bygate, 1998; Zare & Othman, 2015) have expressed concern about the speaking ability of L2 learners.

The Netherlands is not an exception in this regard (Van der Woude et al., 2011, 2012). The speaking skills have been neglected in both L1 (Bonset & Braaksma, 2008; Gelinck, 2000) and L2 classrooms (Fasoglio & Tuin, 2017). In the L2 context, many students believe that the speaking skills are not sufficiently practiced in class (Fasoglio & Tuin, 2017). As a result of this neglect, many secondary school students' speaking skills suffer from deficiencies and the effects thereof become more noticeable and detrimental at the university and college stage (Beeker, 2012). De Vrind (2020) indicated that a shortage of time, large noisy classes, and especially the

absence of effective speaking teaching approaches are among the main factors that prevent doing due justice to the speaking skills in secondary schools.

In an interview conducted with six Dutch secondary school teachers of English about the speaking skills, the teachers revealed that this skill receives the least attention in comparison to other skills in their teaching practice. The interviewed teachers ascribed the neglect of this skill to the fact that they do not have viable teaching tools at their disposal to practice this skill. When asked about using debates, they maintained that debates would be an interesting teaching tool and felt that their students would appreciate them. However, they pointed out that they lack the appropriate pedagogy that would guide them to set up debate activities in their teaching practice and expressed a willingness to receive training about it. Importantly, the neglect of the speaking skills in English classes did not go unnoticed by students. I once asked two classes of *havo 5*[1] to indicate which language skill received the least attention in their English classes from the first form to the fifth grade, and they almost unanimously reported that it was the speaking skill.

In the core objectives set by the Dutch government for secondary schools with regard to the speaking skills of foreign languages, it was specified that pupils should develop the ability to take an active part in discussions and conversations. Debate has been identified as an important means of helping accomplish this objective. Course books in general have not followed these recommendations. In *New Interface (second edition)*, one of the widely used course books for teaching English in the Netherlands, debate activities are seldom employed. For example, in the course books designed for advanced upper school classes (*New Interface vwo 4*, *vwo 5*, and *vwo 6*[2]), debate as a task is used twice (in the course books used in these grades). However, in this task, neither teachers nor students were offered guidelines about how to put these debates into practice. The task asks students to do some research to find arguments that support their sides. It does not specify how to conduct research, and more importantly, it does not provide a practical framework (i.e., debate formats) about how to stage these debates.

In-class debates and writing skills

As our global society increasingly puts a premium on English textually mediated communication, there is an unprecedented need for students to write flexibly, proficiently, and effectively in English. However, Naghdipour (2016) pointed out that 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 et al., 2010), and the effects thereof on

[1] The fifth year of higher general secondary education track.

[2] The fourth, fifth, and sixth year of the preuniversity secondary education track.

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). De Haan and Van Esch (2008) stated that "the Netherlands have no great tradition in the teaching of formal foreign language (FL) writing in secondary education" (p. 7). Naghdipour (2016) and Polio and Park (2016) called for more interventions to inform L2 writing pedagogy about how to help students write effectively.

One of the writing types that poses a great challenge to L2 learners, especially secondary school students, is argumentative writing (Hirvela, 2013, 2017). Instructors experience the task of easing their students into argumentative writing as daunting. 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) maintained that once equipped with the right pedagogy, instructors will be able to make L2 argumentative writing accessible and manageable for L2 learners.

In the Dutch context, students' L2 argumentative writing has received inadequate attention, especially in secondary schools. As a result, Dutch students face many difficulties when composing L2 argumentative essays (De Haan & Van Esch, 2008). In addition, an increasing number of Dutch students take internationally recognized tests, such as Cambridge ESOL, in which English argumentative writing competence is tested. In universities, the majority of students also frequently need to write L2 essays and papers that involve argumentation in one way or another (De Haan & Van Esch, 2005).

Debate pedagogy is one of the innovative pedagogies that may help students foster their L2 speaking and writing abilities. Debate pedagogy may afford learners opportunities to practice and attend to processes that help shape L2 speaking and writing skills. Very few studies in either the L1 or L2 context have examined the effects of debate on language performance in general and on speaking and writing performance in particular. The studies that have explored these effects in L2 so far were based on observations, interviews, and questionnaires (Aclan & Aziz, 2015b; Lustigova, 2011; Zare & Othman, 2015), and most of them were aimed at students in higher education (see Chapter 2).

In-class debates and argumentation skills

As we saw in the previous chapter, argumentation skills are arguably one of the most crucial skills underlying success in everyday life as well as academic and professional contexts (Butt, 2010). Yet, scores of studies have revealed that argumentation skills in the adult population suffer from serious deficiencies and therefore warrant nurturing (e.g., Crowell & Kuhn, 2014; Kuhn, 1991; Kuhn & Udell, 2003). Research has linked deficiencies in argumentation skills to a lack of adequate instruction rather than to an inherent inability of students to acquire such skills

(Butt, 2010; Hirvela, 2017). In the Dutch context, students' argumentative skills have also received inadequate attention, especially in secondary education, despite their manifest importance (Van Eemeren et al., 2015).

In Chapter 2, I pointed out that the debate process is believed to offer an effective pedagogical vehicle for practicing and honing reasoning skills (e.g., Nisbett, 2003; Roy & Macchiette, 2005). However, though there is wide recognition that debate promotes argumentation skills, we still lack an empirical verification of this recognition. Most of the evidence that we have so far is anecdotal, most of which is based on students' perceptions and comes from the L1 context (Scott, 2008). Research investigating L2 argumentative performance is scarce (Qin & Karabacak, 2010).

Taken together, to verify the effects of in-class debates on speaking, writing, and argumentation skills, we need to conduct interventions. However, what hinders this empirical research is the absence of a comprehensive debate task design grounded on meaningful exchange of arguments and that ensures optimal preparation and participation of all students (Brown, 2015). Therefore, the aim of this study was to build a principled and encyclopedic debate task design that responds to, and accommodates, the challenges and particularities of the Dutch secondary school context. This study adopted the EDR approach as this approach has the potential to yield well-balanced interventions (Ivey, 2013; Kwon et al., 2013).

3.3.2 Intended outcomes

In Chapter 2, I demonstrated that the implementation of L2 in-class debates is faced with a number of challenges. These challenges need to be addressed before examining the effectiveness of debates. As Brown and Wilson (2016) contended, debate task design is an influential variable that needs careful attention. Hence, this study set as a goal the development and assessment of a comprehensive debate task design that facilitates meaningful in-class debates and optimizes students' involvement in the learning process. In concrete terms, the objectives of the current study (intended outcomes), which were mainly based on the literature review conducted in Chapter 2, were to:

- design a road map that provides guidelines about how to stage debates in the L2 classroom and that explicates the logistics of debates as they occur in each stage. This is a very fundamental objective because once students are familiarized with the mechanics of debate, debates are expected to be more meaningful and to flow in a natural and logical progression (e.g., Brown, 2015; Omelicheva, 2006; Oros, 2007).
- provide debate formats that involve all students and ensure that all of them have equal opportunities to speak. A debate format is an affective variable that can significantly

shape learning outcomes (Tessier, 2009). Snider and Schnurer (2006) argued that the success of the entire debate process is largely determined by debate formats.

- guide students to carry out adequate preparation. Research (e.g., Brown, 2015; Ericson et al., 2003; Snider & Schnurer, 2006) has underscored the importance of doing sufficient research prior to actual debates. Advance preparation influences students' participation and makes them feel at ease during classroom discussion and participation (Cohen, 1991; Wade, 1994). In addition, preparation activities fulfil the role of reducing the cognitive and linguistic demands on learners (Ellis, 2003).
- create an anxiety-free atmosphere since oral communication in another language can provoke anxiety (e.g., Kormos & Préfontaine, 2017). Because of their competitive and adversarial nature, debates can discourage participation (Bourdillon, 2004; Do & Schallert, 2004; Omelicheva, 2006). Therefore, creating an anxiety-free classroom environment conducive to meaningful debates is of paramount importance. This atmosphere would likely boost students' self-confidence (Garrett et al., 1996) and enhance their willingness to participate in the learning process (Weaver, 2007).

These educational targets were articulated to provide focus for designing the intervention. In the next section, I will seek to identify adequate design principles to attain these objectives.

3.3.3 Literature review

This section presents the literature review that underlies the design principles that informed the construction of the initial prototype of the intervention. Design principles are conceptualized as instruments offering heuristic guidelines for strategy components for obtaining pedagogic effects in classroom (Van den Akker et al., 2006). According to Gravemeijer and Cobb (2006), the literature review should yield a local instruction theory. The local instruction theory should offer conjectures about practical guidelines, leading to the "intended outcomes" of the intervention. Therefore, a literature review was conducted to develop a customizable conceptual framework that would facilitate the implementation of "the intended debates."

The literature reviewed in this study mainly covered two areas of research: debate literature and task-based language teaching (TBLT)[3]. First, I discuss the TBLT literature in Section 3.3.1 in which I first provide a brief overview of TBLT as a teaching approach, discuss the main features of a task, and then argue for the *taskness* of debate. After that, I list the extracted design principles (based on TBLT) that are relevant to the debate task design. Next, I present the design

[3] The literature review also covered one area of content-based instruction (CBI), which seems to be relevant to this study.

principles derived from the debate literature that further guided the construction of the debate task.

Task-based language teaching

Before discussing TBLT, it is important to briefly touch upon communicative language teaching (CLT) since TBLT draws on its principles and teaching philosophy. CLT is an approach that emerged at a time when the existing teaching approaches were felt to be impotent (Richards & Rodgers, 2001). CLT refers to "a diverse set of principles that reflect a communicative view of language and language learning and that can be used to support a wide variety of classroom procedures" (Richards & Rodgers, 2001, p. 105). According to Nunan (2004), CLT is "a broad, philosophical approach to the language curriculum that draws on theory and research in linguistics, anthropology, psychology and sociology" (p. 10). This philosophy has been transformed into various teaching practices (Nunan, 2004; Richards & Rodgers, 2001). CLT approaches teaching from a humanistic perspective in which a premium is placed on interactive processes of communication (Richards & Rodgers, 2001).

The new perspective towards language teaching has had implications for L2 pedagogy. One of these implications is the advent of the TBLT approach (Prabhu, 1987). This approach is premised on the conviction that tasks form a vital basis for language acquisition and hence should be the driving force for syllabus design. Interestingly, during the 1980s the notion of "task" increasingly replaced that of "communicative activity" (Skehan, 2003). It is noteworthy that these pedagogic developments were also accompanied, to a large extent, by a similar underlying claim in the field of second language acquisition (SLA) that language development is conditioned by a naturalistic exposure to, and use, of language (Skehan, 2003).

Tasks emerged as a response to the shortcomings of the traditional approaches, and their position was strengthened by findings from research into how SLA operates in instructed and natural contexts (Long, 1985). Willis (1996) saw TBLT as a logical development of CLT. The TBLT approach is increasingly gaining popularity in the field of SLA, and the study of "tasks" in language learning has connected SLA and language pedagogy since the 1960s (Ellis, 2003). Prabhu (1987) is credited for originating TBLT and learning. He posited that effective learning occurs when students are fully and actively involved in a language task, rather than just learning about language. The fundamental purpose of TBLT is to create a context in which language is communicatively processed by treating it as a tool not as an object (Ellis, 2003). By this, Ellis meant that students should primarily focus on achieving an outcome and not on displaying language when performing a task.

In the absence of actual opportunities to practice L2 outside the classroom, as is usually the case in the L2 context, TBLT provides a unique chance to experience language authentically

through carrying out tasks that mimic how language is utilized in the outside world. Tasks approach language acquisition from a naturalistic perspective. They link the learning that takes place in the classroom to what learners might encounter in their daily life. In other words, tasks provide "motivating communication activities that [bear] some relevance to language use beyond the classroom" (Norris, 2009, p. 580). Widdowson (1998) emphasized that creating an authentic context for language use in classrooms is more important than using the so-called "authentic" materials.

Language in task-based instruction is not taught per se but is treated as a tool for communication (Ellis, 2003; Krahnke, 1987). TBLT inspires learners to forget that they are in a classroom and embrace the idea that they can learn the language through using it rather than directly studying it (Ellis, 2003). Language learners are expected to be more motivated when the focus is on ideas, issues, and opinions than the language itself (Richards & Rodgers, 2001). Meaningful tasks activate learning processes more than form-focused activities and thus offer more opportunities for language learning to move on (Richards & Rodgers, 2001). Moreover, tasks prompt learners to take risks with whatever language they command and to have control over their own learning inside and outside the classroom (Willis, 1998).

Importantly, tasks lend themselves by their nature to evaluation and modification both through microevaluations (i.e., the evaluation of a specific task in a specific instructional context) and macroevaluations (i.e., when whole task-based courses are examined) (Ellis, 1997c, 2012). These processes lie at the heart of EDR, especially in the prototyping and assessment phases, in which different components and aspects of interventions are revised and refined through design-analysis-redesign cycles. This makes EDR as a research method suitable for evaluating and refining task-based interventions.

What is a task? Though TBLT has witnessed an increasing acceptance in the educational world, defining *tasks* has not been easy. Various definitions have tried to capture the distinctive aspects characterizing tasks from different angles. Van den Branden (2016) stated that while these definitions vary, they have a common core: "A task is a goal-oriented activity that people undertake and involves the meaningful use of language" (p. 240). Ellis (2003) pointed out that most definitions address the following dimensions: "(i) the scope of a task; (ii) the perspective from which a task is viewed; (iii) the authenticity of a task; (iv) the linguistic skills required to perform a task; (v) the psychological processes involved in task performance; and (vi) the outcome of a task" (p. 2). On the basis of these dimensions, Ellis (2003) outlined six features of a task:

1. A task is a work plan
2. A task involves a primary focus on meaning

3. A task involves real-world processes of language use
4. A task can involve any of the four language skills
5. A task engages cognitive processes
6. A task has a clearly defined communicative outcome. (Ellis, 2003, pp. 9–10)

Tasks can be "unfocused" or "focused." Unfocused tasks are tasks that are designed to generate opportunities for communicating and using language in general, while focused tasks are tasks that are designed to provide opportunities for focusing on specific linguistic features (mostly grammatical features).

Debate as a task. In this section, I argue that a well-designed in-class debate embodies the distinctive criterial features of tasks outlined above (Ellis, 2003): (1) The work plan specifies what the participants in a task are supposed to do. It is designed to elicit interaction between learners working in groups and pairs. In a formalized debate, students are expected to operate in accordance with a work plan that designates how the activity should take place and that specifies the role of each participant. (2) The primary focus in debate is obviously on meaning, as learners try to convince each other of their viewpoints. (3) While debating, students use their own linguistic resources to defend their viewpoints and to question the validity of the arguments of their opponents. This requires real-world processes of language use. It requires students to convince, prove, disprove, refute, etc., and this occurs in real-world communication. (4) The performance of a well-designed debate requires learners to use up to four skills. It can offer students the opportunity to promote reading through researching the debate topics, listening by listening to each team's opposing arguments, writing as debaters make notes and prepare cases[4], and of course speaking as this skill takes the lion's share of practice. (5) Debates require learners to employ cognitive processes, such as selecting, classifying, ordering, reasoning, refuting, and evaluating information. (6) Finally, debates have a clearly defined communicative outcome; that is, convincing others of a particular viewpoint.

Peng (2016) pointed out that debates "provide a good classroom environment for the implementation of task-based language teaching" (p. 26) because they provide opportunities for information exchange and negotiation of meaning. Nunan (2004) outlined a number of features that characterize real-world tasks. Debate embodies most of these features, which include:

[4] 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.

- defending and arguing a position
- asking peers for their opinions, preferences, and desires
- engaging listeners' attention verbally or nonverbally
- eliciting information and asking clarification questions
- clarifying and restating information as needed
- indicating interests, opinions, or preferences related to class projects
- negotiating solutions to problems, interpersonal misunderstandings, and disputes.

Design principles derived from TBLT

In this section, I list the extracted design principles that emerged from the TBLT literature review. The extracted design principles are of two sorts: macro- and microdesign principles. With macrodesign principles, I refer to principles applying to the whole intervention, and with microdesign principles, I refer to principles applying to specific aspects and activities of the intervention. First, I present the macrodesign principles (D.P.1 and D.P.2), and then I discuss the microdesign principles (D.P.3 up to and including D.P.10).

Macrodesign principles

D.P.1 Granting instructors the freedom to operationalize tasks in a way that fits well with their teaching perspective and context

Given the wide variety of settings and situations in which languages are taught, "it may seem unreasonable … if not downright arrogant, for anyone to propose universally applicable methodological principles for LT" (Long, 2009, p. 375). For this reason, attempts to make recommendations that would apply to all contexts should be precluded (Long, 2009). This applies as well to pedagogical decisions that are constrained by settings and learning context and may be culturally sensitive or dependent. Ellis (2005) elaborated on this point by saying:

> whereas research seeks out general truths, pedagogy must be necessarily contingent and local. Thus, research findings based on one particular context of learning may have little relevance to the particular classroom contexts that individual teachers find themselves working in. (p. 43)

For decades language teaching has been shaped by a number of competing methods, each of which claims to be the "best method" that is able to fare well in every context and learning situation (Nunan, 2004). What these "methods" also share in common is the provision of a package of prescriptions specifying what instructors and learners should do in the language classroom (Richards, 1987). Nunan (2004) rejected this perspective altogether, arguing that:

> it is preferable to identify what works and what does not work through the direct study of the classroom itself. As it is teachers who are the crucial variable in the teaching situation, it is important that teachers should study what goes on in their own classroom. Self-analysis and evaluation will certainly be characteristics of the self-directed teacher. (p. 167)

Because of the nature of the TBLT approach, it is normal and common that tasks are reinterpreted and modified to suit teachers' predispositions, educational beliefs, and established classroom practices (Van den Branden, 2016). East (2012) pointed out that adjustments to the application of TBLT are common and often necessary and that hence teachers should operationalize tasks in a way that makes them appropriate for achieving the set goals and makes them fit better with their teaching style. Also, Richards and Rodgers (2001) maintained that teachers should use approaches and methods flexibly and creatively, transforming and adapting them to mesh with the realities of the classroom. Likewise, Nunan (2004) pointed out that there should be room for teachers to adapt and mold syllabi into an approach that accommodates their objectives, the teaching context, and their students' needs. Van den Branden (2016) correlated a successful implementation of TBLT with taking heed of instructors' educational beliefs and their teaching context. Accordingly, I will provide teachers in other schools with some room to make pedagogical choices that fit their teaching styles (see D.P.7 for some examples).

D.P.2 Attending to the factors that affect students' motivation in task performance

A critical issue for both pedagogy and research is how instructors can configure and operationalize tasks to accommodate learners' affective needs to inspire them to engage thoroughly in task completion emotionally and mentally (e.g., Dörnyei, 2002; Lambert, 2017). Here, a wide range of learner-related variables enter the equation, including (amongst others) cognitive variables (e.g., the learner's memory and brain capacity), socioemotional variables (e.g., the learner's motivation to learn the language and task motivation), and the learner's age and gender (see Dörnyei & Ryan, 2015 for more details). Research has demonstrated that these variables have a potential influence on learners' approach to learning tasks and thus on learning outcomes (e.g., Dörnyei & Ryan, 2015).

It has been consistently demonstrated in the literature on tasks in L2 learning that learner motivation serves a critical role in task outcomes (e.g., Dörnyei & Ryan, 2015; Kang, 2005; Lambert, 2017; Long, 2015; Weaver, 2007). Classroom motivation is a variable state that fluctuates throughout lessons, responding to factors embedded in the design and implementation of tasks (Lambert, 2017). Therefore, it is imperative that teachers become aware of these factors and make adjustments to their tasks accordingly. In designing tasks, the instructor's responsibility is not only to afford ample learning opportunities but also to engage and motivate learners

to put in their best effort (Lambert, 2017). Without the latter, the chances of learning are at stake (Lambert, 2017; Weaver, 2007). In other words, engagement is an indispensable condition for learning (Dörnyei, 2002).

Another important factor that affects students' motivation in task performance is content. When students regard the content under study as interesting, they are likely to learn the language more willingly. Designing tasks that operate on learner-advanced content (for example, with topics that interest learners and appeal to their world of interest) as opposed to teacher-generated content can lead to higher engagement. Lambert et al.'s (2017) study indicated that tasks operating on learner-generated content have more positive effects than those operating on teacher-generated content in all aspects of engagement (behavioral, cognitive, and social components) in L2 use during task performance. Learners feel extremely comfortable and secure when they converse about topics on which they are well informed; these feelings boost their willingness to interact willingly in the conversation (Kang, 2005). Phung (2017) found that L2 learners had negative affective responses about tasks revolving around unfamiliar topics, whereas they were more positive towards tasks involving familiar topics.

Another piece of evidence that further supports this design principle comes from CBI. The CBI, which takes content as a point of departure or as an organizing principle (Richards & Rodgers, 2001), provides further evidence that the content under study is an affective variable and should therefore be selected with utmost care. The CBI sees that one way to safeguard the selection of appropriate and relevant content is through involving learners in this process (Brinton et al., 1989). This perspective seems to be relevant to our intervention development and will be implemented through allowing students to select the debate topics they consider to be relevant. Relevant content is expected to "increase motivation in the language course, and thus to promote more effective learning" (Brinton et al., 1989, p. 3).

This design principle will be further operationalized in our debate task design through taking heed of the factors that can possibly affect students' task motivation and performance. This includes, for example, employing the debate formats that students find engaging and make them feel at ease.

Microdesign principles derived from TBLT

In general, microdesign principles concern task procedures (i.e., the methodological procedures followed when performing a task). Task procedures can be best investigated in the phases in which tasks are performed. Though different designs have been proposed in TBLT research (e.g., Ellis, 2003; Long, 1985; Prabhu, 1987; Skehan, 1998; Willis, 1996), they all have three principal phases in common: the "pre-task" phase, which involves the activities that teachers and students can undertake before they perform the task; the "during task" phase, which is the

main part and concerns the task itself; and the "post-task" phase, which concerns the procedures that follow task performance. There are a number of task procedure variables that can affect students' engagement and performance (e.g., Ellis, 2003; Lambert, 2017; Long, 2015; Skehan, 1998; Van den Branden, 2016). I reviewed the task procedures that are relevant to debate task design as they occur in each stage.

Pre-task stage. During the pre-task phase, teachers and students prepare the task performance organizationally, cognitively, and socioemotionally (Van den Branden, 2016). From an organizational perspective, teachers need to provide their students with clear instructions on how the task should be performed and what is expected of them. From a cognitive point of view, the topic at hand and relevant vocabulary need to be introduced as well as activating learners' prior knowledge. From a socioemotional angle, teachers need to address learners' questions and provide a safe environment for taking risks with output.

Ellis (2003) pointed out that the pre-task stage enables teachers to create the conditions that enable tasks to facilitate acquisition. Skehan (1998) noted that these conditions play the role of introducing new language that learners can turn to when completing the task, marshaling their own linguistic resources, alleviating processing load, and stimulating them to undertake tasks in more demanding ways. Norris (2009) referred to this stage as a *task input* phase. He emphasized its importance, as it exposes learners to rich input that establishes linkages to the contexts in which the main task occurs. He also added that exposure to input during this phase can induce a considerable amount of L2 acquisition. Weaver (2007) stated that having the opportunity to preplan and/or rehearse what is intended to be said during a task is also likely to boost learners' perceived communicative competence and hence bolster their willingness to communicate. In what follows, the pre-task methodological options that are relevant to our debate intervention are discussed.

D.P.3 Providing a model

Ellis (2003) proposed giving students the opportunity to observe a model of how a task can or should be undertaken. This can be done in the form of presenting a text (oral or written) demonstrating how the task can be ideally performed. Activities raising learners' consciousness about specific features of the task performance can also accompany the presentation of the model (Ellis, 2003). This includes, for example, the strategies that can be employed to surmount the potential problems that can arise in the course of communication, or the steps that can be followed to convince "the other" of one's standpoint during a debate. Both Skehan (1996) and Willis (1996) maintained that simply observing others perform a task is likely to reduce the cognitive load on the learner.

D.P.4 Employing nontask preparation activities

Students can also perform a number of activities in the pre-task phase that can boost performance and ease the burden of the main task. These activities thus serve the role of reducing the cognitive or linguistic demands that can encumber learners (Ellis, 2003). They enable them to map the different angles and aspects of the task and provide the equipment necessary to perform it. Nontask activities can include, for example, reading articles, watching documentaries, and vocabulary tasks (see Willis, 1996 for other activities). Undertaking them would likely make learners more knowledgeable about the topic area of a task and provide them with useful vocabulary they can employ in the during-task phase.

D.P.5 Facilitating task repetition

It is commonly acknowledged that learners improve their L2 performance when repeating the same or similar tasks (Qiu & Lo, 2017). Task repetition can concern the repetition of the same procedure with different content (procedural repetition) (Qiu & Lo, 2017) or repeating the same task with the same content (Ellis, 2009). Skehan (2014) pointed out that repeated performance capitalizes upon traces of the partial lemma retrieval that has occurred during the first performance. In task repetition, the first performance of the task functions as preparation for (or as a pre-task activity before) further performances (Ellis, 2005). Task repetition makes retrieval of prior content easy, and through "freed up processing capacity," learners are better prepared to monitor or attend to form (Moser, 2012). In other words, by repeating the same or similar tasks, learners "'buy time' not only to do mental work on what they are about to communicate but also to access and (re)formulate words and grammatical structures more efficiently, effectively, and accurately" (Ahmadian, 2012, p. 1).

D.P.6 Encouraging and facilitating pre-task planning

Preplanning can be operationalized through a number of methodological options. One of these options concerns whether to provide guidance to students on what to plan. Ellis (2003) noted that students prioritize content over form when they are left to decide themselves what to plan. However, when guiding them in what to plan, teachers may channel their students' focus into form or content, or into both of them. The guidance can be detailed, involving instructions related to planning content and language, or undetailed. Having the opportunity to preplan and/or rehearse what is intended to be said in a task is also likely to boost learners' perceived communicative competence and hence bolster their willingness to communicate (Weaver, 2007).

During-task phase. Ellis (2003) noted that teachers have two methodological options at their disposal in the during-task phase: task performance options and process options. Task performance options concern options that can be planned before the performance of a task and that sketch how the task is to be performed. Whereas performance options can be elected prior to the actual performance of the task, process options are taken while the task is being completed. These options involve making decisions on the spot while students are performing tasks.

D.P.7 Implementing focus on form through corrective feedback

The rationale underlying focus on form rests on the premise that L2 learning is most effectively developed if learners are afforded the opportunity to direct their attention to form while involved in meaning-focused language use (Ellis, 2012). Long (2015) stated that linguistic problems should be treated as they arise. This approach, in his eyes, is in line with SLA research and is in accord with the assumption that learning grammar develops from language use. Long (1991) defined focus on form as follows: "Focus on form overtly draws students' attention to linguistic elements as they arise incidentally in lessons whose overriding focus is on meaning or communication" (pp. 45–46). Focus on form is included in this section because it can function as a process option, but as will be discussed below, it can be embedded in all phases.

Ellis (2003, 2012) distinguished between two kinds of focus on form: planned focus on form and incidental focus on form. Planned focus on form takes place in focused tasks in which the target linguistic features are predetermined. Incidental focus on form relates to the use of unfocused tasks in which linguistic forms are treated as they arise during task performance. Nevertheless, students and teachers may choose to pay attention to various forms while carrying out the task (Ellis, 2003, 2012). In such a case, attention to form will be extensive (as different forms being treated briefly) rather than intensive.

It should be noted that focus on form in the TBLT context principally differs from focus on forms occurring in traditional lessons. One of the main differences is pointed out by Prabhu (1987), who considered correction during a task as incidental rather than systematic. In incidental correction, only "tokens" are discussed. Learners see it as "a part of getting on with the activity in hand, not as a separate objective" (Prabhu, 1987, p. 63).

The leading scholars of TBLT like Ellis, Long, and Skehan regard focus on form as an essential part of TBLT. It constitutes one of the main ways of addressing grammar in TBLT (Ellis, 2009). SLA researchers also consider it an essential requirement for L2 acquisition, especially when it comes to adult learners (Ellis, 2005). However, they disagree with regard to when it should be addressed. Willis (1996) noted that it should occur in the pre-task and post-task phases but not in the main-task phase. Long (1985, 1991) advocated its introduction only in the main-task phase. Skehan (1998) noted that it should be mainly attended to in the pre-

task phase. However, Ellis (2003) took a different position. He suggested that focus on form can occur in all phases of the task, and he further maintained that this need not conflict with a primary focus on meaning (especially in the during-task phase). We adopted Ellis's perspective because of its flexibility and compatibility with D.P.1. An important way of implementing focus on form is through corrective feedback (Ellis, 2018).

Post-task phase. Ellis (2003) noted that a number of options are available in this phase. One option that seems to be relevant to debate task design concerns focus on forms activities:

D.P.8 Employing focus on forms activities in post-task phase

Once students are done with the task, their attention can be directed to focus on forms. Ellis (2003) argued that this focus does not imperil the "taskness" of the task. Ellis (2003) maintained that two obvious methodological questions need to be dealt with when attending to form in the post-task phase. The first relates to which forms teachers should address. Ellis (2003) stated that teachers should particularly elect forms that their students used incorrectly while completing the task to fill gaps in their L2 knowledge. As for how many such forms a teacher should discuss, Ellis said that teachers have two approaches at their disposal: treating a single form intensively or a number of forms extensively. The second methodological question concerns how the forms in question should be addressed. Ellis (2003) proposed four options: review of learners' errors, consciousness-raising tasks, production practice activities, and noticing activities. The first option gives teachers the opportunity to note down errors (with their context) while students are performing the task in groups. In the post-task stage, teachers can analyze these errors together with their students and also provide both a brief explicit instruction and an opportunity to practice and use the forms in question correctly. Consciousness-raising tasks can be used as independent tasks in their own right or as follow-up tasks to explicitly channel students' attention into the forms they failed to use correctly. In these tasks students need to get guidelines to help and encourage them to identify errors in their utterances, repair them, and figure out an explanation.

Focus on forms will be dealt with in the post-debate stage. We will attempt to employ the activities that best suit the recipient (students). It is important to note that the instructors in other schools will be granted the freedom to employ the activities that are in consonance with their pedagogical perspective and teaching style.

Debate literature and design principles

To further distill flexible and rigorous principles that would enable the intended objectives to be realized, I investigated the debate literature. The debate literature is expected to provide more

specific guidelines about the construction of the intervention. However, very few academic works have been identified that described how debates were staged in class. In addition, most of these studies were conducted in universities and covered competitive debate. More importantly, the pedagogical principles underlying the debate activities were premised on anecdotal and personal experience rather than on theoretical grounds and scientific evidence (see Chapter 2.7 for more details). On the whole, our inspection of the debate literature led to extracting two new relevant design principles (D.P.9 and D.P.10) and further provided extra support and information about the operationalization of two design principles (selection of appropriate debate topics and debate preparation) discussed in the previous section.

D.P.9 Selecting debate formats that have a clear structure and that ensure equal involvement of all students

For a debate to be successful, there should be a debate format with a clear structure that sets out the order and time limits of the debate components (e.g., constructive speech, rebuttal, etc.) (Bibby, 2014). A debate format is an affective variable that can have a significant effect on learning outcomes (Tessier, 2009). This is why Tessier (2009) advised instructors to "think carefully about the design of a debate to maximize the learning potential" (p. 146). Snider and Schnurer (2006) even contended that much of the success of the entire debate process banks on debate formats.

A debate format should provide a clear framework of how debates should run (Snider & Schnurer, 2006). It is certainly of importance that students should know from the outset when to take turns to present, for example, their constructive speeches and rebuttals. Being instructed about the structure and rules that govern debates will enable students to stage them with confidence, reassurance, and enthusiasm. Such an atmosphere is unlikely to have a marked bearing on the quality of debates and will optimize the objectives and benefits that are to be reaped from them. In effect, any debate format has two sides: an affirmative side and a negative side. The job of the affirmative side is to defend and advocate the adoption of a resolution, while the opposition side opposes and refutes it.

When selecting debate formats, it is imperative to take into account class dynamics and the desired learning outcomes (Alén et al., 2015). As Snider and Schnurer (2006) argued, there is no right or wrong way of staging debates. Yet, debate formats should accommodate students' abilities and give them the same opportunities to speak (Rybold, 2006; Snider & Schnurer, 2006). In the debate literature different debate formats have been suggested, most of which consist of one to four debaters on each side.

In principle, there is no right or wrong way of having debates (Snider & Schnurer, 2006). In Chapter 2, we saw a number of in-class debate formats that were proposed in the debate

literature. However, the majority of these formats came from the competitive L1 debate context or were modeled on them. There is a scarcity of debate formats specifically designed for the L2 context (e.g., Lieb, 2007; Stewart & Pleisch, 1998; Zare & Othman, 2013).

The debate formats below were selected because: (1) they seem to suit our context; (2) they are expected to ensure equal involvement of as many students as possible; and (3) they can smoothen the path for a natural flow of debates:

- Two-team debate (see Snider & Schnurer, 2006)
- Policy debate (see Rybold, 2006)
- Public forum debate (see Rybold, 2006)
- Lincoln-Douglas debate (see Rybold, 2006)
- Spontaneous Argumentation (see Snider & Schnurer, 2006)

These debate formats go through the following stages: constructive speeches, rebuttals, and cross-examination.

Constructive speech. During constructive speeches, each speaker/team makes a case for their side by presenting cogent evidence. While constructive speeches are presented, opponents should just listen and make notes of the arguments being advanced as this is crucial for the next stage.

Cross-examination. Cross-examination is a period of questioning that happens mostly at the end of constructive speeches (this depends on the debate format) in which one team asks their opponents questions. The purpose of the cross-examination is to gather information to support one's case and to highlight inadequacies in the opponent's case.

Rebuttal. Rebuttal is used as an umbrella word that includes attempts to refute arguments through pointing out flaws and inadequacies in them as well as attempts to reestablish the validity of debaters' position through countering attacks being launched against their line of reasoning (Driscoll & Zompetti, 2003). Opponents during this stage should just listen and make notes. If they have got questions or requests for further illustrations, they should note them down and present them during the cross-examination stage.

D.P.10 Grading debates to prompt students to take them more seriously

Grading debates is essential to drive students to take them more seriously (Oros, 2007; Scannapieco, 1997). It goes without saying that students generally tend to mobilize their effort and perform to the best of their capacity during activities for which they receive a mark. Scannapieco (1997) remarked in this regard that "the inherent determination of most students

to obtain a good grade usually motivates them to prepare well for this activity [debate]" (pp. 957–958). In a similar vein, Omelicheva and Avdeyeva (2008) pointed out that students are typically more motivated to mobilize their intellectual energy in tackling challenging problems when their work is graded.

Oros (2007) warned that instructors should not underestimate the challenge of simultaneously managing the timing of debates, listening to the arguments advanced, evaluating different skills, and eventually recording all this into decipherable notes. Therefore, he advised developing a "score-sheet" in advance and noting down comments and scores immediately or soon after class as the debate is mostly fresh in one's mind.

In addition to extracting two new design principles, the debate literature review (Chapter 2) guided us to extract two supplementary design principles that concern D.P.2 and 4.

D.P.2 Attending to the factors that affect students' motivation in task performance

The first design principle that will be supplemented in light of debate literature concerns attending to the factors that affect students' motivation in task performance. Employing content that is relevant in students' eyes is one of these very important factors. Both TBLT and CBI research have provided robust evidence indicating that designing tasks that operate on content considered relevant by the learner is likely to generate higher engagement, motivation, and performance. Cognitive research has also evidenced that attending to and incorporating what interests students boosts the learning process; by contrast, when "ignoring the personal world of the learner, educators actually inhibit the effective functioning of the brain" (Caine & Caine, 1991, p. 86). In a similar vein, research around debate has revealed that staging debates around relevant topics is a critical condition that needs to be met for debates to be effective (e.g., Alford & Surdu, 2002; Rybold, 2006; Zare & Othman, 2015). Research has shown that topic interest is an affective variable that enhances learners' involvement in the learning process and positively impacts their language development (Lee & Pulido, 2017). Therefore, it is important to facilitate the selection of topics that students consider relevant and can relate to, as this can even shape students' attitude towards the whole debate process (Dörnyei, 2002; Dörnyei & Kormos, 2000). Zare and Othman (2015) advised instructors to involve learners in the process of selecting debate topics to enhance participation during debates. On the basis of these findings, D.P.2 will be supplemented by:

D.P.2a Facilitating the selection of interesting and engaging debate topics

D.P.4 Employing nontask preparation activities

As has been argued in the previous chapter, sufficient advance preparation is key to a successful and meaningful debate. One research source that has been widely used in the debate preparation stage is newspaper and magazine articles (e.g., Brown, 2009; Snider & Schnurer, 2006). These kinds of articles are mostly easily accessible through the Internet and are informative. They contain relevant information and are not challenging to read (Snider & Schnurer, 2006; Stewart, 2003). Reading articles also helps students to distinguish between primary and secondary information and between trivial and important facts (Iberri-Shea, 2013). Therefore, it seems that there is merit in requiring students to use these kinds of articles as a source of preparation. Hence, D.P.4 will be supplemented by the following design principle:

D.P.4a Requiring students to read articles during the preparation stage

3.4 Prototyping stage

The literature review has resulted in identifying 10 principles that I believe are viable and robust to guide the development of a comprehensive debate intervention. I will piece these design principles together in an attempt to form a cohesive task design that scaffolds the process of debating. On the whole, the design principles seem to encapsulate all the essential pedagogical characteristics of a comprehensive debate intervention, including procedures, activities, and context.

The prototyping stage enables theory to be linked to practice through operationalizing the findings of the literature review into a practical framework conjecturing how debate activities should run in different phases. In short, this stage concretizes the findings of the previous stage by embedding the emerged design principles into the initial prototype of the debate design. The initial prototype is expected to be usable (expected practicability) and effective (expected effectiveness) in achieving the desired outcomes for the target context. To demonstrate the actual practicality of the intervention, we need to field-test it (Nieveen, 2007; Plomp, 2007).

As has been outlined above in Section 3.3, there are diverse activities that can be performed in each stage. Our selection of these activities is informed by research, our experience, and the practicality of the teaching context for which the intervention will be developed.

Table 3.1 displays the derived design principles, in which stage they will be put into practice, and the goals they are meant to fulfill. D.P.1 and D.P.2 are general design principles and do not concern specific aspects of the debate task design.

Table 3.1 Summary of the extracted design principles

Stages and activities	Design principles	Rationale (in order to …)
	Macrodesign principles	
	D.P.1 Granting instructors the freedom to operationalize tasks in a way that fits well with their teaching perspective and context	Tailor the intervention to instructors' teaching style and perspective
	D.P.2 Attending to the factors that affect students' motivation in task performance	Increase students' motivation to try their best during task completion
	Microdesign principles	
Preparatory stage		
PowerPoint presentation	D.P.3 Providing a model	Inform students of the debate project and acquaint them with the mechanics of debate
Selecting topics	D.P.2a Facilitating the selection of interesting and engaging debate topics	Select interesting and engaging debate topics
Pre-debate stage		
Reading articles	D.P.4 Employing nontask preparation activities D.P.4a Requiring students to read articles during the preparation stage	Boost performance in the during-task phase Enable students to become knowledgeable about the debate topic and search for relevant arguments
Rehearsing cases	D.P.5 Facilitating task repetition	Boost speech delivery
During-debate stage		
Pre-task planning	D.P.6 Encouraging and facilitating pre-task planning	Boost performance in the during-task phase
Debate formats	D.P.9 Selecting debate formats that have a clear structure and that ensure equal involvement of all students	Facilitate smooth and engaging debates
Focus on form	D.P.7 Implementing focus on form through corrective feedback	Attend to students' accuracy
Grading debates	D.P.10 Grading debates to prompt students to take them more seriously	Stimulate students to take debates seriously
Post-debate stage		
Focus on forms	D.P.8 Employing focus on forms activities in post-task phase	Attend to students' accuracy

3.4.1 The initial prototype

In this section, I outline the initial debate prototype based on the extracted design principles. The activities of the prototype are presented in the order in which they are performed.

Preparatory stage

Prior to the start of debates, students should be informed about the mechanics of the whole project through a PowerPoint presentation. This presentation elucidates how debates should run and what students have to do in each stage. This discussion will not be extensive since many of the debating skills are natural communication practices that are used daily (Snider & Schnurer, 2006). The debate formats will be elucidated so that students know in which stage, for example, they have to present their constructive speeches and rebuttals and hence pre-plan their performance (see D.P.6). In line with the suggestions of Ellis (2003) and Willis (1996), the debaters will be presented with a model (video) demonstrating how debates can be staged (see D.P.3).

The next step involves selecting topics. As has been argued in D.P.2a, selecting appropriate topics is essential to increase students' engagement and make debate a successful learning activity. To guarantee that appropriate topics, which engross students and pique their curiosity, are selected, students will be involved in selecting them. They will be required to submit topics that are meaningful and have fair ground to debate on both sides. The submitted topics will be gathered on a list and submitted to the vote. The topics that receive the majority of votes will be on the final list. Also, topics that relate to current events will be proposed and selected in consultation with students. Attention will also be paid to the formulation of the "qualified topics." I will attempt to frame them in straightforward words, so that no room is left for ambiguity. In other words, the topics must be clear and accurate in all students' minds. After topic selection, the next step concerns assigning students to positive and negative teams. I will allow students to defend the side they are convinced of. Students are expected to voice preference for one side more than the other. To level the ground for a fair contest, some students (mostly higher-achievers) will be requested to play devil's advocate.

Pre-debate stage

As has been evidenced by research, proper preparation plays an invaluable role in facilitating a successful task completion (see D.P.4 (a)). This stage also fulfills the role of counteracting classroom apprehension and enhancing classroom participation. In-class debates can often be intimidating for many students. Speaking up in front of a class is a huge challenge that can encumber students. Many students are simply not comfortable as speakers. They also lack confidence in their grasp of L2 or even in presenting good arguments. They are afraid of being laughed at or being confronted. Wade (1994) argued that advance preparation is among the central factors that influence students' participation and make them feel at ease during

classroom discussions. Cohen (1991) also considered preparation a major source that lowers the threshold for participation in the classroom.

Weir (2009) cautioned that instructors should never assume that students will come to debates prepared. To address this, Oros (2007) proposed grading debates, as "students generally will not prepare significantly for a debate for which they will not be graded" (p. 297). To further accentuate the importance of the preparation phase, preparation will be one of the components on the basis of which the debaters' marks will be decided. To make sure that students carry out proper and sufficient research, we will require them to read articles (see D.P.4a). Monitoring students' preparation phase is essential since this phase is critical for a successful and effective debate. Not doing this would ask for situations in which students arrive at class unprepared and unmotivated.

During debate

In D.P.9, five debate formats have been selected. To meet the goal of involving all students in each debate session, the time limits for each speaking turn (e.g., constructive speeches, rebuttals, etc.) will be reduced to one minute. Stewart (2003) argued that shorter time limits facilitate more equal opportunities to participate in debates and decrease the amount of information that a debater needs to communicate, thereby alleviating the burden on less competent debaters and listeners to not be overwhelmed by too much information.

While constructive speeches are presented, opponents should just listen and make notes of the arguments being advanced, as this is crucial for the next stages. Scannapieco (1997) pointed out that the management of nondebating groups is of critical importance for enhancing the potential of debate for the whole class. So to ensure that the nondebating students in other teams actively listen to debates, they will be required to make notes of the debaters' language use and do some form of peer grading. Furthermore, they will be tasked, in addition to listening critically to the advanced arguments, with noting down any grammatical mistake they spot and any word they learn from each other. The students will be urged to rely on the context to guess at the meaning of the new words. In line with D.P.7, the mistakes that will be made by students (unplanned focus on form) will be occasionally addressed during debates but not by interrupting the flow of communication. This will be done at the end of debate stages. The erroneous utterances will be discussed briefly during this stage.

Students and instructors alike should realize that leveraging the potential of debate is conditioned by adequate preparation. Adequate preparation manifests itself, in the first place, in the quality and quantity of constructive speeches being delivered. Just presenting flimsy arguments and employing a fallacious line of reasoning in a short time obviously proves that

the preparation is not adequate. As mentioned before, preparation will receive credit in the assessment rubric.

Post-debate stage

After staging debates, it is important to provide students with some feedback. This feedback will primarily focus on language use, including grammar and vocabulary. Ellis (2003) underlined that focusing on forms in the post phase does not jeopardize "the 'taskness' of the task." Aclan and Aziz (2015b) also advocate the inclusion of this stage in debates to raise students' awareness of their mistakes and hence strike a balance between fluency and accuracy. Instructors can discuss the provided feedback at the beginning of the next class after debates. In the post-debate stage, errors will be tackled more intensively and extensively. Errors will be discussed without embarrassing the students who made them. Moreover, we will stimulate students to advance errors they spotted in the during-stage for discussion.

It is important to note that the instructors in the other schools (participating in this research) will be offered the opportunity to address errors (during and after debates) in the way that works best for their groups and resonates with their teaching philosophy (see D.P.1).

The role of the instructor

The teacher has several responsibilities during each debate stage (i.e., before, during, and after debates). Before debates, the teacher needs to spell out the debate process to students and emphasize the fact that debating will be used as a learning experience and not to yield "winners" and "losers." This will likely assuage students' uneasiness with speaking in a group. The teacher also needs to delineate the rules and debate formats to be used for debates and facilitate the selection of appropriate debate topics.

In the during-debate stage, the instructor serves the role of observer and facilitator. Learning will not be effective when the instructor regulates or controls everything that occurs in the classroom and assumes the major responsibility in the learning process, marginalizing the learner's role (Worthen & Pack, 1992). On the other hand, when learning is student-centered and the instructor oversees the process, this will likely help lower the stakes for student participation. Students feel more at ease when interacting with their classmates than with teachers (Byrd, 2008). Willis (1992) even noted that "in the absence of the teacher, [students'] interaction becomes far richer" (p. 180). Garrett et al. (1996) pointed out that during debate, the instructor also has the role of setting the educational environment and the emotional climate. The educational environment concerns arranging debates in a way that debaters can see and hear each other comfortably, while the emotional climate should put students at ease by reinforcing a cooperative, nonthreatening, and nonadversarial climate. Also, the instructor

needs to stay alert when the task is being undertaken in order to be able to identify erroneous utterances and address them adequately.

Assessment

Another role that instructors need to fulfill during debates is assessing debaters' performance (D.P.10). As has been argued by Oros (2007) and Scannapieco (1997), grading debates induces students to take debates more seriously. To perform this task properly given that the instructor has other roles to fulfill, we will assess a limited number of debaters in one debate session to be able to accommodate the other demands of debate. Yet, we will not inform students about when their performance will be assessed. This way students will always come well prepared and put in a solid performance in all debates. However, we will inform students about how their performance will be assessed. The following criteria that cover components that measure content and oral proficiency aspects will be employed, drawing on Brown (2004) and Oros (2007): *constructive speech*, *rebuttal*, *fluency*, *vocabulary*, *grammar*, and *preparation*. Each component will be rated on a five-point scale, with lower scores indicating poorer performance.

3.4.2 Cycles of implementation findings and implications

After outlining the initial prototype of the intervention following the principles of EDR, I will move to the next phase, which consisted of iterative research microcycles during which the initial prototype was put to the test. This phase pursued the goal of constructing a comprehensive debate task design embodying the prespecified goals in Section 3.2. Accordingly, prototypes 1–3 of the debate task design were developed, implemented, and formatively evaluated. Each cycle of design research consisted of the implementation of a prototype of the debate task. Each prototype was formatively evaluated, leading to a further cycle of refinement and development resulting in a new prototype. Formative evaluation is pivotal to the prototyping phase. Its results are meant to locate shortcomings in the design and provide suggestions about how to address these shortcomings. The formative evaluation questions and focus were related to the extent to which the intervention achieved the intended outcomes and how students responded to the task design and modifications. Different tools (see McKenney & Reeves, 2012), which included students' small-group discussions, observation notes, interviews, and questionnaires regarding debating formats, were used as part of the data analysis that helped to identify the factors that improved or hindered the intervention realizing "the intended objectives." The confluence of these data sources informed the modifications to the intervention. The cycle of implementation involved three classes: vwo 3, vwo 4, and havo 5. Experience and research have shown that these classes are able to stage decent debates and are also able to provide a reliable assessment of the learning process (e.g., Van Strijen & Morino, 2015).

In general, the modifications and refinements of the intervention were guided by the criteria characterizing high-quality interventions proposed by Nieveen (1999): relevance, consistency, practicality, and effectiveness (see Section 2).

First cycle

Implementation. Following Plomp (2007), we focused particularly on consistency and practicality in the first implementation of the debate intervention. That is, we examined whether the components of the debate task design were logically and smoothly connected to each other and whether the task design is usable and workable. The first implementation involved five debates in two classes (vwo 3 and vwo 4).

Method. The participants of the analyses discussed in this cycle were two Dutch classes (vwo 3 and vwo 4) consisting of 49 students (aged 14–16) in a secondary school in Rotterdam. The vwo 3 class consisted of 26 students (14 males and 12 females), and the vwo 4 class comprised 23 students (10 males and 13 females). Both groups received three sessions of English a week. For the purpose of this study, one session was devoted to debate. Both groups were engaged in five debates, which were conducted in accordance with the initial prototype outlined above. The debates revolved around controversial topics selected by the students, and in each debate a new debate format was introduced. The students were informed beforehand of the debate topic and the side they would defend so they could do the necessary preparation. We were granted permission by the participants' parents to use their children's data for research purposes.

Observation data were gathered during debates, and after the debates more data were elicited through group discussions. The observations focused primarily on the coherence of the components of the debate task design and the extent to which they were usable. In the group discussions, the students were prompted to reflect upon and comment on all stages and components of the debate task design.

Findings. Based on the classroom observations and group discussion data, we concluded that the debates were motivating for almost all students. The students from both groups indicated that they enjoyed the learning experience offered by the debates. However, our observations and student discussions identified a number of flaws that hindered the optimal practicability and consistency of the debate task:

- The structure of the debate formats seemed to be complex and confusing
- Not all students did sufficient preparation
- Some students found it challenging to find articles relevant to the topic of debate

- Some students brought articles to the class, but still we felt that they had not read them
- Constructive speeches were short

Implications. Overall, the above findings prompted further refinement to counter the shortcomings that emerged from the first implementation. By examining the observed debates and scrutinizing the group discussions, we saw ways to ameliorate the emerging problems and further tweak the debate task design to make it more consistent and usable for students. The modifications that were made to the debate task design concerned debate formats, preparation, and debate forms (see Appendix A).

Debate formats. As has been argued in D.P.9, using appropriate debate formats is a key factor that can have a huge impact on the whole debate process. Omelicheva (2006) contended that the structure of debate formats can be adjusted to meet specific educational goals. Randolph (2007) reported that she adjusted the debate format she used each year on the basis of the feedback of her students. From the first implementation of the prototype, we learned that the employed debate formats were structurally complex. Therefore, we decided to simplify the debate formats in consultation with the students (through group discussion). Involving the students in designing the new formats was expected to result in accessible and engaging debate formats and is in line with the philosophy of D.P.2, which underscores the importance of attending to the factors conducive to students' engagement in the learning process. The new debate formats consist of three stages: constructive speech, rebuttal, and clash. After each stage, students receive a break to prepare for the subsequent stage. In the new debate formats, clash, which takes place at the end of debates, replaces cross-examination. In clash, teams further challenge each other's perspectives by exposing each other's inconsistencies in reasoning as well as reinforcing their stands with additional evidence. In what follows, the new debate formats will be delineated.

Debate Format 1. The class is split up into two teams (see Table 3.2). One team is affirmative, while the other is negative. Students are free to choose which side they want to defend. To ensure that everyone has an equal opportunity to speak, cards are used. Each student receives three cards. Each time a student talks, s/he hands in one card. On the first occasion, speakers from both teams have to deliver their constructive speeches in one minute. On the second and third occasions, they have to present their rebuttals (30 seconds on each occasion).

Table 3.2 Debate Format 1

Constructive speech	1 minute
Team A: First speaker: Affirmative constructive speech	1 minute
Team B: First speaker: Negative constructive speech	1 minute
Team A: N^{th} speaker: Affirmative constructive speech	1 minute
Team B: N^{th} speaker: Negative constructive speech	1 minute
Rebuttal	
Team A: First speaker: First affirmative rebuttal	30 seconds
Team B: First speaker: First negative rebuttal	30 seconds
Team A: N^{th} speaker: First affirmative rebuttal	30 seconds
Team B: N^{th} speaker: First negative rebuttal	30 seconds
Team A: First speaker: Second affirmative rebuttal	30 seconds
Team B: First speaker: Second negative rebuttal	30 seconds
Team A: N^{th} speaker: Second affirmative rebuttal	30 seconds
Team B: N^{th} speaker: Second negative rebuttal	30 seconds

Debate Format 2. This debate format involves two teams (affirmative and negative) of two speakers (see Table 3.3). The debate starts with each speaker holding a constructive speech of one minute. After that, both teams receive a short break to prepare a rebuttal, which they present together in approximately two minutes. The debate ends with the speakers facing each other in the clash.

Table 3.3 Debate Format 2

Constructive speech	
Team A: First speaker: Affirmative constructive speech	1 minute
Team B: First speaker: Negative constructive speech	1 minute
Team A: Second speaker: Affirmative constructive speech	1 minute
Team B: Second speaker: Negative constructive speech	1 minute
Preparation	10 minutes
Rebuttal	
Team A: Both speakers: Affirmative rebuttal	2 minutes
Team B: Both speakers: Negative rebuttal	2 minutes
Clash	5 minutes

Debate Format 3. This debate format involves two debaters (see Table 3.4). One debater is on the affirmative side and the other one on the negative side. Both debaters receive an opportunity to present their arguments and rebuttals, and face each other in the clash. Students receive one minute in the first two rounds and five minutes in the third round, in addition to five minutes preparation time between the constructive speech and rebuttal.

Table 3.4 Debate Format 3

Constructive speech	
First speaker: Affirmative constructive speech	1 minute
Second speaker: Negative constructive speech	1 minute
Preparation	5 minutes
Rebuttal	
First speaker: Rebuttal	1 minute
Second speaker: Rebuttal	1 minute
Clash	5 minutes

Preparation. During preparation interval, each speaker/team (this depends on the format being used) receives an opportunity to regroup their thoughts and prepare for the next stages. Their main focus should be on identifying flaws in opponents' arguments presented during the constructive speech stage and on how to respond to these flaws. When debaters work in teams (Debate Format 2), they should coordinate their efforts and also confer about the strategy they need to use to present their rebuttals. Intervals are also meant to prepare for the last stage, clash. Students could prepare requests for further information and other questions that could expose weaknesses in their opponents' reasoning.

How to operationalize the formats. Now I will elucidate how Debate Formats 2 and 3 can be operationalized.

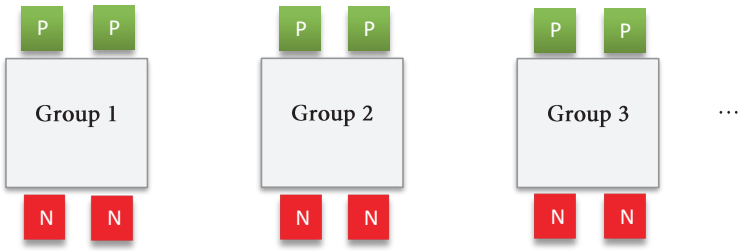


Figure 3.1 Seating arrangement of students in Debate Format 2

With regard to Debate Format 2 (see Figure 3.1), the debate commences with asking the debaters of Group 1 to deliver their constructive speeches alternately. After delivering the constructive speeches, this group will be offered a preparation interval of 10 minutes to prepare for the rebuttal and clash stages. While the first group prepares for the subsequent stages, the instructor asks the second and then the third group to present their constructive speeches and subsequently grants them 10 minutes each to prepare for the next stages. After that, the instructor gets back to the first group and asks its debaters to do the rebuttals. In this way, the

instructor alternates between constructive speeches and rebuttals until all groups have carried out these two stages. Finally, all groups will be engaged in the clash stage at the same time. Asking each group to do the clash separately is unfeasible because of time constraints. During this stage, the instructor will move from group to group, basically just to listen and make sure that the debate is on the right track. The same procedures of this methodological option can also be applied to Debate Format 3. It is important to note that students are not allowed to interrupt each other in the first two stages and are required to participate in each stage.

Sufficient preparation. Since conducting enough research is one of the keystones that facilitates successful debate performance, guiding students to perform this activity properly is of paramount importance. In the first prototype, the students were required to do preparation, but we noticed that many students did not do it properly. As a result, most constructive speeches (first stage of debate) were short and contained flimsy arguments. Also, some students reported that they were unable to find appropriate articles relevant to the debate topic. Moreover, it seemed that some of the students who brought articles to the class had not read them. To address these shortcomings, we felt that we needed to revise D.P.4a and to add a supplementary design principle (D.P.4b) to help students prepare optimally for actual debates. Accordingly, we reformulated D.P.4a as:

D.P.4a Requiring students to read articles and summarize them during the preparation stage

And added:

D.P.4b Requiring students to write cases during the preparation stage

In effect, in the next implementation:

- We gave the students one article providing arguments for both sides of the topic, and we tasked them to find at least one more article. Students were asked to print out copies of the articles so that they could mark relevant arguments and easily refer to them throughout the debate.
- We required the students to make short summaries of the articles to make sure that they had read them.
- We required the students to write their cases using their own words and submit them at the end of each debate. The cases were reviewed and returned.
- We encouraged the students to make use of notecards to help them to cope with the nerves and to produce a smooth constructive speech.

Second cycle

Implementation. In our implementation of the revised design, we concentrated our evaluation on whether the modifications found an answer to the weaknesses that surfaced in the first implementation and how the students responded to the new changes. In addition to *consistency* and *practicality*, we evaluated the *effectiveness* quality of the intervention as well (see Nieveen, 1999; Plomp, 2007). The second prototype was also implemented in the same groups of the previous cycle, with each group taking part in five debates. The debates followed the same procedures as in the first implementation. In addition to classroom observations and group discussions, data were also collected through a five-point scale questionnaire about the new debate formats. The questionnaire consisted of eight items (two items per debate format). It addressed the extent to which the students liked the new debate formats and would like to participate in debates involving these debate formats.

Findings. Our observation of the debates and group discussions suggested that the new changes seemed to get us closer to the intended outcomes. The modifications to the previous prototype appeared to result in better debates in terms of students' engagement and performance. The measures that were taken to ensure better preparation were successful. This manifested itself, for example, in the fact that the students managed to produce longer constructive speeches.

With regard to debate formats, the students felt more at ease with the new simplified debate formats and were able to carry out more meaningful debates. We believe that the new debate formats, which ensure that every student has an equal opportunity to speak, enable debates to march forward in a simple, natural, and logical progression. As Table 3.5 shows, the mean score for all debate formats is above 3 (the neutral point). They range from 3.77 to 3.90. This indicates that the students liked the new debate formats and expressed preparedness to participate in debates orchestrated by these formats.

Table 3.5 Descriptive statistics about students' attitude towards the new debate formats

Debate formats	*M* (range 1–5)	*SD*
Debate Format 1	3.78	1.30
Debate Format 2	3.77	1.06
Debate Format 3	3.90	.91

However, though a substantial improvement was secured after the first implementation, we further learned from observations and interviews that:

- Though the constructive speeches improved in terms of length, they were fragmentary and not coherent.

- The students came with long lists of words, but the majority failed to actively use them.
- It was noticed that it was too demanding for the students to simultaneously rate their classmates, note down their arguments, and attend to their language use (through noting down the words they learn from them as well as the errors they make).
- The students found the experience of staging debates in front of the class very frightening and a barrier to good performance.

Implications. After pondering on the findings above, we made the following modifications:

- We provided the students with a list of the most common cohesive markers and asked them to use at least three of them in their cases.
- We asked the students to employ the new learned words in their cases as far as possible.
- We relieved the students of the burden of rating their classmates. So they were asked to solely focus on arguments and language use (see the debate form in Appendix A).
- It was apparent that a number of students felt uneasy about debating in front of their classmates. We noticed this even after the implementation of the first prototype, but we underestimated this burden. After further discussions with the students, we decided not to require them to stage debates in front of the class.

These modifications required the adjustment of D.P.4b, which we reformulated as follows[5]:

D.P.4b Requiring students to write cases in which they should use cohesive markers and the new learned words (when possible) during the preparation stage

Third cycle

Implementation. After modifying our design in light of the outcomes of the previous cycle, we conducted debates in the same groups using the same number of debates and following the same procedures. Data were collected as in the previous iterations through classroom observations and group discussions. We focused our attention on verifying the extent to which our modifications produced the intended outcomes.

Findings. The classroom observations and group discussions indicated that the adaptations made significant improvements to the debate task design, resulting in improvements in student engagement. The students reported that they found the modified debate task more facilitative of meaningful debates. We also observed that freeing the students of the task of rating each other resulted in locating more mistakes in each other's performance, in addition to learning

[5] see Appendix B for an overview of the final list of the design principles underlying the developed debate task design.

more words (from each other). Overall, the last refinements provided satisfaction that the intervention had been improved substantially and that we were getting closer to the complete realization of the desired outcomes. Design-analysis-redesign cycles allowed us to revise different components and aspects of the debate task. Though we felt at this stage that we were near to realizing the desired outcomes, still we felt that one point needs further refinement. In the previous implementations, each student had to listen to all constructive speeches and rebuttals of their classmates. Though this proved to be beneficial since the students get a lot of input from which they profit, particularly at the argumentative, grammatical, and lexical level, it was a bit worrisome and even tough to listen actively to all these inputs. Additionally, with a class size exceeding 24 students, it is not feasible to involve all students in each debate. Therefore, devising an extra methodological approach that responds to these concerns (accommodating larger classes and offering a methodological variety) is more than warranted.

Implications. In response to these limitations, we added another implementation approach to Debate Formats 2 and 3 that involves a simultaneous presentation of constructive speeches, rebuttals, and clash. If Debate Format 3 is used, the instructor can ask the speakers of all affirmative teams to deliver their constructive speeches simultaneously. Then, the speakers of the negative teams follow suit. If Debate Format 2 is employed, the speakers of each team should alternate simultaneously in presenting their constructive speeches. The instructor travels from group to group, monitoring the process by keeping time and making sure, for example, that opponents listen attentively and make the necessary notes. It is important to note that this approach should take place in a spacious classroom so that debaters do not disrupt each other. After all debaters have presented their constructive speeches, a preparation interval of 10 minutes should be granted. Then, each speaker/team of each group takes turn in staging rebuttals, again simultaneously. Lastly, all groups engage in the clash.

The main advantage of this approach is that it is time saving and can be implemented in large classes. Time is saved as constructive speeches, rebuttals, and clash are presented at the same time. However, the main drawback is that each group of debaters works separately. They only talk and listen to their group. Yet, the time-saving advantage dilutes this drawback to a certain extent. Once students are done with the clash, there will be some time left. This time can, for example, be used for asking some groups to reconduct the three stages of debate (or one of them) while requiring other students to listen attentively and make notes about language use.

Up until this stage, the development of the debate task design has involved vwo 3 and vwo 4 classes, while the interventions in the next chapters (Chapters 4–9) also involve havo 5. Also, the iterations in each cycle involved five debates, while our interventions involve 10 debates.

McKenney and Reeves (2012) commented that "it does not make much sense to endeavor to ascertain the effectiveness of an intervention as a whole when it has only been partially implemented" (p. 139). Because of this, we felt that it was important to pilot the whole intervention with 10 debates in havo 5.

3.5 Final pilot stage (assessment phase)

Method. The participants in this cycle comprised 20 Dutch students studying English at the B1–B2 level in a secondary school in Rotterdam. The class consisted of 20 students (6 males and 14 females) and received three sessions of English a week. For the purpose of this study, one session was devoted to debate. The debates were conducted in accordance with the revised procedures laid out in the last cycle. In addition to observation notes and group discussions, formative evaluation drew on three semi-structured interviews involving nine students.

Findings. The final cycle incorporated the learning experience from all the previous cycles in order to examine the functioning of the intervention in its totality. It therefore consisted of a semi-summative evaluation (Plomp, 2007). In general, the students were positive about the debate task design. They reported that the developed debate task provided the opportunity to have meaningful debates. They pointed out that the task design created a pleasant environment that lowered the threshold of participation. They also remarked that the components of the debate task design formed a coherent whole. Our observations concurred with the students' perspectives.

During the interviews, the participants were also offered the chance to voice their criticism of the debate task. The main criticism leveled at the debate task was about the labor-intensive preparation the debates entailed. Many students maintained that preparing for a debate was arduous and the least appreciated aspect of the debate process. However, a number of participants pointed out that the preparation time dramatically decreased in the last debates as this participant declared:

> [the preparation of] the first debate lesson lasted really long, two hours or so, and the last one about half an hour.

Though time-consuming, a number of participants emphasized the importance and utility of preparing well for debates.

Importantly, it should be noted that the majority of the interviewees considered the "debate homework" less daunting than the homework of other subjects. They see more value in the "debate homework" because it involves variety (e.g., using the Internet), and it primes them to

perform well in a contest they look forward to. Nevertheless, homework will always be experienced as a burden; therefore, it is important, if space allows, to give students the opportunity, at least occasionally, to make the necessary preparation in class.

All components of the intervention were assessed (through group discussions, classroom observations, and interviews) for practicality and effectiveness, including, for example, debate formats, debate forms, topics, etc. Different aspects of components were rated high by the students, including clarity, the importance of the topics, accessibility, and ease of use. The EDR-inspired cycles thus established that all the quality criteria, namely relevance (content validity), consistency (construct validity), the practicality of the intervention (e.g., ease of use of the different components of the debate task design such as debate formats), and perceived effectiveness were met. In sum, we believe that the objectives (see Section 3.2) that guided this study have been fulfilled. Hence, we can conclude that the quality of the learning process is good enough to merit moving to the next stage.

The emerged debate task design is not of a rigid character. Many of its components lend themselves to modifications to mesh with different teaching perspectives and context realities. It is true that the proposed debate task design was developed in one school. However, the debate task design was later evaluated along with two teachers from two other schools that welcomed the intervention. The evaluation mainly focused on two questions: (1) to what extent has the debate task design realized the objectives it was set to achieve?; and (2) to what extent does the debate task design satisfy the quality criteria specified by Nieveen (1999), i.e., relevance, consistency, practicality, and effectiveness? With regard to the first question, the teachers thought that the debate task design successfully met the set goals. They reported that the developed task design provides clear guidelines about how to stage meaningful debates, that the debate formats are accessible and ensure that all students have equal opportunities to debate, that students get clear guidance about how to perform sufficient preparation, and that the debate atmosphere created by the debate task design seems to be anxiety free. As to the second question, the teachers were of the opinion that the debate task design is relevant, implementable, and coherent in terms of its activities, and they expected it to be effective.

3.6 The role of the instructor as an educational research designer

The researcher has held multiple roles in this design research study. In particular, I developed the intervention, field-tested its prototypes, and prepared most of the learning materials. As an insider researcher, I cannot claim that the research process was fully unbiased due to my proximity to, and involvement in, the development of the intervention. This proximity and involvement might have caused biased interpretation, analysis, and reporting (see Jönsson & Lukka,

2007). Keeping the distance needed for an unbiased perspective can be quite difficult in design research (Kouwenhoven, 2013). However, being an insider research designer allowed me to intensively experience and examine the educational problem at hand from the perspective of an instructor who has been using debates for many years in teaching practice. It is true that in studies where "a researcher is intimately involved in the conceptualization, design, development, implementation and researching of a pedagogical approach, ensuring that researchers make credible and trustworthy assertions can be a challenge" (Barab & Squire, 2004, p. 10). Nevertheless, I attempted to be conscientious and not permit my role to influence the research results. Furthermore, the fact that formative evaluation drew on diverse sources of data elicitation restricted any influence I could exert on the research process as an insider researcher. Barab and Squire (2004) pointed out that despite the potential threats being associated with being a participant research, researchers should not "remain detached from the research context, [they] are implored to intervene where possible, using interventions as opportunities to examine core theoretical issues and explore learning" (p. 10).

3.7 Reflection

The use of EDR was highly effective for the design and adaptation of the debate task design. The EDR approach provided the opportunity to design, implement, and evaluate various prototypes, which gradually started to approximate the objectives of the study. The main strength of EDR lies in the design-analysis-redesign cycle, which allowed us to create the conditions for realizing the intended outcomes. This flexible iterative design process enabled the design process to remain responsive to formative evaluation emerging from each cycle. The design research approach resulted in a flexible debate task design that provides step-by-step guidance on how to conduct debates in and beyond the context in which they were designed.

It is possible and even justifiable to assume that other teachers (in other schools) might want to adjust certain components of the debate task design to fit their teaching style and context. The good news is that the debate task design we have developed seems to be adaptive and customizable, and this makes it stand a chance of being successfully implemented in different schools. An example of the flexibility of the design concerns how to approach focus on form. For example, some instructors may prefer to provide immediate feedback after each constructive speech, while others might rather wait until the completion of the whole debate. Another example concerns debate formats. It is possible that a particular instructor may see fit to employ one particular debate format more often than another. Hence, this degree of flexibility is needed to ensure that teachers do not sacrifice a particular perspective or preference they are convinced of, as this can be detrimental to this promising teaching tool.

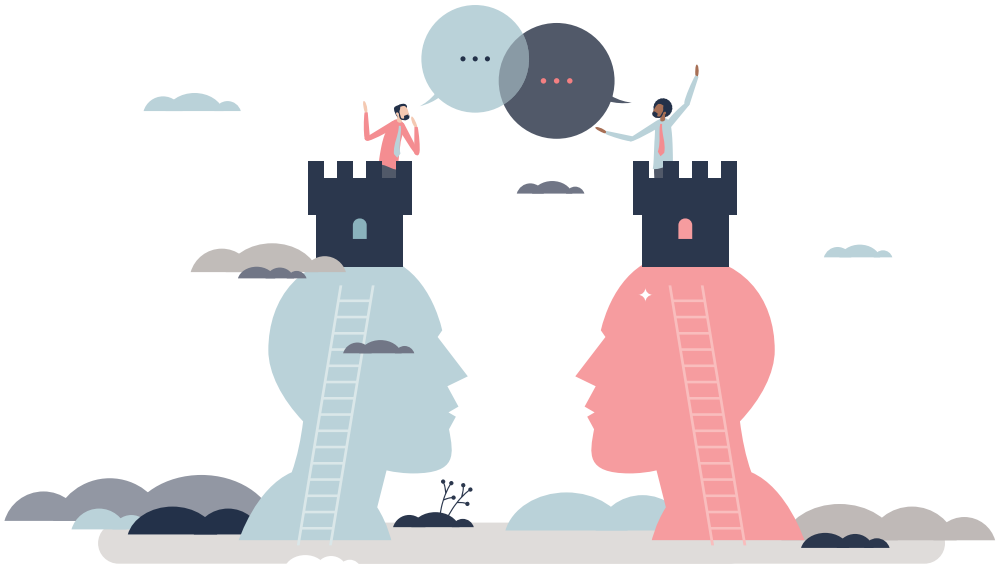
3.8 Conclusion

The literature is replete with assertions and indications suggesting that in-class debates hold potential for enhancing students' speaking, writing, and argumentation skills. To empirically test the credibility of these claims, it behooves us to develop an intervention that does justice to the full potential of in-class debates. Engineering the debate intervention from the EDR perspective allowed us to work along a clear road map that provided clear directions that facilitated working towards the "intended outcomes." The design research philosophy of EDR allowed us to view students as partners and collaborators in the design process. This enabled us to elicit their perspectives, which yielded illuminating insights that further fed and strengthened the intervention.

Organizing and staging debates may seem demanding. However, when we have a comprehensive and field-tested road map in which each stage is laid out and supported by practical and feasible steps, the challenge fades away gradually. This study has helped us to construct a balanced debate task design that is robust enough to empirically test its potential for improving students' speaking, writing, and argumentation skills. Importantly, the design is adaptive in nature, lending itself to implementation beyond the local context in which it was developed. This will enable us to test it in different schools to increase the external validity of the ensuing results.

Chapter 4

Invest in What Energizes Students to Learn: Investigating Students' Attitude towards Debate in the Second Language Classroom



This chapter is based upon: El Majidi, A., de Graaff, R., & Janssen, D. (2015). Invest in what energizes students to learn: Investigating students' attitude towards debate in the foreign language classroom. *Journal of Language Teaching and Research,* 6(5), 924–932. https://doi.org/10.17507/jltr.0605.03

Abstract

Debate has noticeably been penetrating our educational practice over the last few decades. Numerous studies have revealed many benefits of debate, including sharpening up debaters' critical and analytical thinking. Debate has also been deemed an effective pedagogical tool for language development. Its effectiveness has been mainly ascribed to its ability to integrate the four language skills, stimulate learners to be active, and engage them in interactive activities that entail working collaboratively to negotiate meaning. In many studies, students ac-claimed debate as fun and enjoyable. This study investigated the factors that underlie the 44 participants' (belonging to two classes) positive attitude towards debate as an instructional tool in the second language classroom. Our data consisted of a questionnaire and interviews. The study has revealed that the factors that positively shape the participants' favorable attitude towards debate are *active participation*, *challenge*, *teamwork*, *fun*, *critical thinking*, *language proficiency*, and *debate vs. course book*. The independent samples t-test showed that both groups extended the same relevance to these factors in shaping their attitude towards debate, with the exception of the critical thinking factor. However, a comparison between the male and female participants revealed a number of significant differences.

Keywords: debate, attitude, foreign language teaching, pedagogical tool

4.1 Introduction

Debate is an inherent part of our life. We are constantly involved in attempts to convince others and to influence their views and decisions. Debate occurs everywhere—at home, at school, and in meetings. Debate can be formal, as in parliaments, or informal, such as between friends.

Debate enhances debaters' communication skills. In the context of the increasing importance of communicating knowledge effectively and in that of social and political debates, it is imperative to have strong communication skills (Akerman & Neale, 2011). Studies have shown that people with strong communication skills find their way quickly to leadership roles and promotion at work (Snider, 2008). Debating skills are even extolled as indispensable ingredients for success in all walks of life. In other words, "debating creates the skills you need for success wherever your life may lead you" (Snider, 2008, p. xiv).

Debate is an important tool for an enriching learning experience (Lieb, 2007), and using it as a teaching approach brings a lot of benefits to learners (Zare & Othman, 2013). It offers teachers the chance to engage their students in a variety of activities that inspire students to explain, justify, convince, and counter. Also, debate not only helps students deepen their comprehension of the issue/topic in question and foster their critical thinking abilities, but it also helps them enhance their language proficiency (Zare & Othman, 2013). Moreover, debate enables teachers to involve students in an engaging and cooperative learning process that facilitates the interaction of students with each other and with the content as well.

Students enjoy debating (Kennedy, 2009). Many studies have reported students' liking for debate and their preference for this pedagogical tool over other tools (see the literature review). It goes without saying that students lend their support to every activity that they find enjoyable. What is enjoyable fires students with enthusiasm and makes them love learning. Not surprisingly, the activities that students experience as fun bring the best out of them. Moreover, what is fun holds students' attention and safeguards it from distraction.

This study tries to uncover the factors that make second language (L2) students enjoy debate. The theoretical framework that underlies the hypothesis of this study stems from the literature and a small-scale exploratory qualitative study.

4.2 Literature review

Debate is "the process of inquiry and advocacy, a way of arriving at a reasoned judgment on a proposition" (Freeley & Steinberg, 2005, p. 4). Debating involves a process of considering different viewpoints and making a judgment (Goodwin, 2003; Kennedy, 2007, 2009). Debate can function as a performance as well as a method that conveys ideas and arguments; it is a communication event in which the mode of operation can be oral or written (Snider &

Schnurer, 2006). Snider and Schnurer (2006) stated that the use of debate as a teaching method dates back to ancient Greek and Roman philosophers. They mentioned that Confucius and other Chinese philosophers wrote famous treatises that characterized debate as a valuable method of learning.

In-class debate promotes critical thinking (Rashtchi & Sadraeimanesh, 2011). Critical thinking is "thinking about how you think" (Rybold, 2006, p. 74). In our modern times, critical thinking is a necessity since we are continuously snowed under with information. Worthen and Pack (1992) maintained that the ability to evaluate information critically is a must-have for every person. They added that when students are stimulated to think critically, they will be better prepared to cope with the future and its complexities. Moreover, Rashtchi and Sadraeimanesh (2011) noted that "practicing critical thinking changes learners from passive receivers of new materials into critical thinkers" (p. 386).

Debate-like activities can enhance disciplinary learning. Goodwin (2003) said that both teaching experience and empirical research have proved that debate helps students develop content mastery. Bellon (2000) argued that if students are not given a chance to debate about important concepts they are presented with in class, they will not be able to "develop deep or mature understandings of course content" (p. 172).

Students gain a lot of benefits when teachers use instructional strategies that invite active engagement (Doody & Condon, 2012). Active learning is any activity that engages students in a classroom other than listening passively to an instructor (Faust & Paulson, 1998). Bellon (2000) pointed out that cognitive research has shown that successful classrooms are interactive, whereas students learn less when forced into passive roles or practices. Fallahi and Haney (2007) regarded debate as "an excellent form of active learning" (p. 83). In addition, debate helps students to cultivate an open-minded acceptance of various views on a given topic (Kennedy, 2007).

In-class debate benefits the whole class and not only the enthusiastic and excellent students. This benefit extends even to passive students (Stewart & Pleisch, 1998). Stewart and Pleisch (1998) reported that their teaching practice revealed that passive students enjoy working on language tasks, undertaking research, and writing papers for debates. Warner and Bruschke (2001) compared this benefit that touches every student in the debating process to a gym class. In a gym class, all students benefit from some exposure to physical fitness.

In-class debate involves students in a learning process that involves practicing the four major skills of language (Alasmari & Ahmed, 2013; Lieb, 2007; Rybold, 2006; Snider & Schnurer, 2006; Zare & Othman, 2013). Zare and Othman (2013) contended that in-class debate facilitates access to linguistic input and output. In addition to speaking, debate can provide L2 students with an opportunity to promote critical reading through researching the

topics of debate, critical listening by listening to each team's opposing arguments, and writing as debaters prepare cases[1] and make notes. However, the debate community has not produced enough research that has demonstrated that participating in debate cultivates the debaters' language proficiency (Omelicheva & Avdeyeva, 2008). The very few works that have been published on debating for L2 learners have mainly focused on the format and procedures of debating (Zare & Othman, 2013).

Debate empowers students to intellectually challenge and outperform each other in their arguments and the way they frame them. Debaters challenge and push each other to the limit. In other words, "the competitive process pushes students to excel" (Rowland, 1995, p. 108). Amid such competition, students sharpen up their reasoning abilities and language proficiency (Lieb, 2007).

Furthermore, the environment created in debate is conducive to collaborative learning. In debate, students work collaboratively to resolve the issues raised by the debate topics. By so doing, students foster communicative and cooperative skills and discern the importance of conjoint effort in creating successful learning. In their study, Fallahi and Haney (2007) reported that "a total of 80% experienced a feeling of group accomplishment or teamwork during the debate, and 64% preferred working with a team rather than working alone" (p. 86).

Importantly, students support and value the use of debate as a teaching method (Alford & Surdu, 2002; Kennedy, 2009). Stewart and Pleisch (1998) reported that their students consistently approved of debate, with a rate above 80%, as "the best/most interesting course activity" in their course surveys between 1994 and 1998. They also pointed out that not even one student in four years recommended excluding debate from the course. Khan et al. (2012) conducted a study involving students of health economics in which the participants were divided into two groups of active and nonactive debaters. The participants' perception of debate as a teaching tool was evaluated with a structured questionnaire before and after the debates. In addition to the increase in their knowledge of the topics being discussed, the active debaters also reported an increase in their interest in debate. In another study, Kennedy (2009) pointed out that the majority of the participants in five debates spoke favorably of these debates and even mentioned that they would consider using debate as an instructional tool. In a similar vein, a number of other studies reported that the students who took part in debates described them as fun and educational (Alford & Surdu, 2002; Fallahi & Haney, 2007; Omelicheva & Avdeyeva, 2008; Van de Woude et al., 2011).

[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.

Also, the teachers who have employed debate in class hold a positive attitude towards this tool. For example, Jiménez et al. (2011) reported that the teachers who experimented with debate in the industrial classes were positive about this tool. They characterized the experience as motivating and rewarding.

Yet, students' positive attitude towards debate has not been extensively researched. It is important to gain a comprehensive insight into the underlying factors that generate and sustain this positive attitude. Park et al. (2011) stated in this regard that "… further research is needed on this teaching strategy, not only in terms of faculty perceptions and experiences but also in terms of student perceptions and experiences" (p. 14).

Hill (1982) pioneered the study of what motivates students to debate. In his study, in which he used a questionnaire, he asked debaters to list in order of importance the reasons that accurately describe their motivation to be involved in debate. From this study, six categories emerged: *educational*, *social*, *competitive*, *career preparation*, *financial*, and *miscellaneous*. Wood and Rowland-Morin (1989) replicated the study of Hill, using a five-point Likert scale questionnaire. Jones (1994) pointed out that both Hill's and Wood and Rowland-Morin's studies were restricted by the methodology, in that the categories of the former were based on truncated and nondeveloped answers, while those of the latter were based on a Likert scale, which did not allow for an in-depth response that accounted for the importance of the given reasons. With these methodological restrictions in mind, Jones (1994) replicated the study of Hill in an attempt to accurately locate the specific reasons for the motivation behind debaters' behavior. This methodological refinement and study replication are crucial in Jones' eyes for justifying debate as a worthwhile activity. Jones adopted a qualitative approach by using interviews and field research as a source of data. Ninety-eight debaters were observed and interviewed at six intercollegiate debate tournaments. After analyzing and coding the content, five primary categories—*cerebral*, *competition*, *heuristic*, *social*, and *miscellaneous*—in addition to one secondary category—*intellectual reinforcement*—emerged.

The studies of Hill (1982), Wood and Rowland-Morin (1989), and Jones (1994) explored only the factors that account for debaters' liking for debate in the first language. These studies, moreover, did not investigate the differences in the perceptions of males and females. The current study, therefore, endeavors to dig up these factors in the L2 context and investigate how they are conceptualized by male and female students since a number of studies have shown that gender is a relevant variable in the perception of L2 learning. For example, Aldosari (2014) concluded that male and female students differ in their motivation and attitude towards L2 learning. Siebert (2003) also found a number of significant differences in beliefs among male and female students with regard to L2 learning and strategy use. This study set out to answer the following research questions:

RQ.1 What are the factors that shape L2 students' attitude towards debate?

RQ.2 How are these factors perceived by male and female students?

4.3 Method

4.3.1 Participants

The participants of this study comprised 44 Dutch students belonging to two classes at a secondary school in Rotterdam: a lower secondary class group (Grade 9) consisting of 25 students, including 12 males and 13 females (aged 14–16), and an upper secondary class group (Grade 11) consisting of 19 students, including 7 males and 12 females (aged 16–19). The first group was studying English at the B1 level and the second group at the B2 level. Both groups received three English sessions of 50 minutes a week from the first author. During this study, which involved 10 debates, one session was dedicated to debate, and in the other two sessions the participants received regular lessons based on a course book.

4.3.2 Intervention

The participants were required to submit a list of controversial topics that interested them. Each student was required to hand in two topics. The topics were gathered, ordered into a list, and submitted to a vote. The students were asked to mark the topics they would eagerly like to debate. Accordingly, a final list was made. This list included topics like: abortion should be banned; students should wear uniforms, etc.

The students were informed about the topics that would be debated at least one week in advance so that they could do the necessary preparation, and they were asked to research both sides of each issue. During debates, the students who were not debating were asked to make notes of their classmates' performance. We used three debate formats: splitting the class into two teams (one team in favor and the other against), debating in groups of four debaters (two students in favor and two against), and a one-to-one debating format.

4.3.3 Data collection

To map the underlying factors that motivate students to debate, we developed a questionnaire (Appendix C) and conducted interviews to enhance confidence in the ensuing findings. In this connection, Griffee (2012) argued that "questionnaires represent a data collection process that is considered 'a mile wide and an inch deep,' as opposed to interview data which might be described as 'an inch wide and a mile deep'" (p. 139).

To generate and specify the critical concepts that should be addressed by the questionnaire, a small-scale exploratory qualitative study was carried out in the form of a series of focus

group and one-to-one interviews. "Such a design is effective in improving the content representation of the survey and thus the internal validity of the study" (Dörnyei & Taguchi, 2009, p. 110). It is noteworthy that a number of the questionnaire items were modeled on the *attitude*, *fun*, *challenge*, and *critical thinking* scales of Jones (1994) and Gardner (1985).

The questionnaire was piloted with a group of 18 students, which was similar to the target sample. The obtained data were submitted to item analysis using SPSS to check whether the internal consistency of each construct was larger than the .70 threshold (see Dörnyei & Taguchi, 2009). The internal reliability of the constructs ranged from .74 to .85. This means that the items in each construct were internally consistent. This guaranteed that the questionnaire did not contain any glitches that could jeopardize the quality of the questionnaire in the final administration. It is important to note that after the piloting phase, the factor of *language proficiency* emerged as a potential factor that could positively influence the participants' attitude towards debate. This factor was included in the final questionnaire.

To make sure the developed questionnaire was straightforward and easy to understand with the least possible cognitive effort, nine students of similar ages to the participants were recruited to go through its items. The items of the questionnaire were randomly ordered to avoid frustrating the respondents with repetitive content. In addition, the wording of some items was reversed to prevent bias response.

4.3.4 Procedure

The questionnaire, which consisted of 40 items, addressed eight constructs of the underlying motivating factors to debate (see Table 4.1). The participants were asked to rate each item on a five-point Likert scale: 1 = *strongly disagree*, 2 = *disagree*, 3 = *neutral*, 4 = *agree*, and 5 = *strongly agree*.

For the interview, 10 students (eight females and two males) agreed to participate. A semi-structured format was used. To make the participants feel at ease and tackle the issue of status of inequality between the first author (teacher) and the participants, a group interview was opted for. The participants were interviewed in groups of two or three in a classroom. The interviews lasted approximately 20 minutes. They ended once the researcher felt all aspects of the questions of interest had been exhausted. At the beginning of the interview, the participants were informed of the purpose of the research and interview.

Table 4.1 Definition of the eight constructs that may influence students' attitude towards debate

Constructs	Definitions
Attitude	Attitude towards debate
Fun	Pleasure gained from debating
Active participation	Active involvement in the learning process rather than passively absorbing information
Critical thinking	Thinking critically to find strong arguments to outperform the opposing team-/debater and locate flaws in their arguments
Challenge	The challenge posed by debating and by convincing classmates
Teamwork	Working collaboratively during debates
Language proficiency	Influence of debate on language proficiency
Debate vs. course book	The attractiveness of debate as opposed to working with course books

4.4 Results

4.4.1 Analysis of the questionnaire

Table 4.2 The internal consistency of the constructs

Constructs	Number of items	Cronbach's alpha
Attitude	5	.80
Fun	5	.80
Active participation	5	.85
Critical thinking	5	.83
Challenge	5	.70
Teamwork	5	.88
Debate vs. course book	5	.97
Language proficiency	4	.76

Table 4.2 illustrates the internal reliability of the scales, which was calculated through the Cronbach's alpha coefficient. The internal consistency met expectations on all scales with the exception of the language proficiency scale, which did not reach the .70 threshold (remember this scale was not piloted). Some fine-tuning was then needed. With the deletion of item 22, the internal consistency rose to .76. This was the only modification necessary since the internal consistency of the other scales exceeded the threshold, ranging from .70 to .97. This signifies that the items of each scale are strongly interrelated and thus measure the same underlying constructs.

Table 4.3 Descriptive statistics

Constructs	Grade	*n*	*M*	*SD*
Attitude	Grade 9	25	3.97	.66
	Grade 11	19	3.77	.53
Fun	Grade 9	25	4.06	.62
	Grade 11	19	3.90	.42
Active participation	Grade 9	25	4.01	.70
	Grade 11	19	3.64	.60
Critical thinking	Grade 9	25	4.08	.68
	Grade 11	19	3.50	.61
Challenge	Grade 9	25	3.93	.74
	Grade 11	19	3.65	.42
Teamwork	Grade 9	25	3.82	.67
	Grade 11	19	3.61	.83
Debate vs. course book	Grade 9	25	4.48	.73
	Grade 11	19	4.07	.87
Language proficiency	Grade 9	25	3.53	.74
	Grade 11	19	3.65	.70

As Table 4.3 shows, the mean score for all constructs is above 3, ranging from 3.50 to 4.48. This indicates that the participants perceived all the constructs as playing a role in generating the positive attitude they hold towards debate. This fact is espoused by a one-sample *t*-test, which showed that all mean scores on all scales deviated significantly from the scale center. This means that the participants acknowledged the importance of all factors in instilling a favorable attitude into them towards debate.

As Table 4.3 also reveals, the *course book vs. debate* factor was rated as the most influential factor in both groups, with an average mean of 4.48 in Grade 9 and 4.07 in Grade 11. The *fun* factor was also rated high in both groups. It was rated as the third most influential factor in Grade 9 and as second in Grade 11. Strikingly, the *critical thinking* factor was rated as the second most influential factor in Grade 9, while it was rated as the least influential factor in Grade 11.

The language proficiency scale was not rated high by both groups. Item 36 of this scale, which states *I like debate because it is important for my writing skill*, was the lowest-rated item. Its deletion would substantially raise the mean score of the scale. Prior to debates, students were asked to hand in a debate preparation form (as homework), in which they had to write, among other things, a case. Getting a writing task as homework is not what students generally like. Because of this, the students could have associated this item with homework and rated it low, thereby affecting the mean score of the whole scale. The mean scores of the other constructs

are very close. This means that the participants attached the same importance to these factors in influencing their attitude towards debate.

To compare the scores of Grades 9 and 11 for each construct, an independent-samples test was conducted. As Table 4.4 demonstrates, only the *critical thinking* construct reached significance. This means that there is no statistically significant difference in the means of other constructs. In other words, the respondents in both groups more or less attached the same importance to these constructs.

Table 4.4 Independent-samples test results

Constructs	*t*	*df*	*p* (two-tailed)
Attitude	1.06	42	.294
Fun	.95	42	.346
Active participation	1.84	42	.072
Critical thinking	2.86	42	.007
Challenge	1.48	42	.146
Teamwork	.93	42	.355
Course book	1.66	42	.103
Language proficiency	-.57	42	.566

To compare males' and females' perceptions of the factors in question, an independent-samples *t*-test was conducted. This test showed that females found debate more fun than males: M(males) = 3.8 (SD .62), M(females) = 4.2 (SD .42); $t(42) = -2.5$, $p < .05$. Also, females attached more importance to the *active participation*: M(males) = 3.6 (SD .65), M(females) = 4.0 (SD .65); $t(42) = -2.6$, $p < .05$ and *challenge*: M(males) = 3.6 (SD .70), M(females) = 4.0 (SD .53); $t(42) = 2.2$, $p < .05$ factors than males. Interestingly, males extended more importance to *critical thinking* than females M(males) = 4.0 (SD .82), M(females) = 3.6 (SD .57); $t(42) = -1.9$, $p < .05$.

4.4.2 Analysis of the interviews

To read between the lines of the questionnaire and dig deep into it, interviews with 10 participants were conducted. In these interviews, the participants were confronted with the potential factors that may shape their attitude towards debate.

Debate is fun. All the participants in the interviews extolled debates, stating that they liked them and would certainly participate in them once the opportunity arose. Interestingly, some participants even advocated the use of debate in other subjects, like German, Dutch, and biology. However, some argued that debating in French and German was not sensible at that moment because their current command of these languages was not good enough to stage adequate debates. One participant even expressed her willingness to join a debate club and participate in debate competitions.

Active participation. Debates may create fertile ground for active engagement through providing students with the opportunity to explain, clarify, analyze, and rebut. The participants appreciated this quality of debate. One participant said that she liked debate because it facilitated active participation in the learning process. It was also pointed out that debate was preferred because

> everyone is stimulated to participate [and that] everyone is engaged and gets something out of it.

Critical thinking. In the interviews, all the participants praised debate for pushing them to think critically and analytically. Two participants stated in this regard:

> It is pleasant to think critically to prove why your arguments are OK, and the ones of your opponents aren't.

> I like debate because it makes you reflect.

Challenge. Debate may pose an interesting challenge to students as one participant contended. That challenge is a source of enjoyment was also corroborated by the study of Williams (2006), in which he concluded that "[students] liked being challenged and they thought that challenge did lead to enhanced learning and enjoyment" (pp. 9–10). Williams (2006) also concluded that "to really be challenging [teachers] needed to include higher-order thinking skills, for example a need to evaluate or justify a statement or action or to manipulate information before applying knowledge to a problem" (p. 8). Interestingly, some participants also reported enjoying the challenge of trying to outshine their classmates in the way they framed arguments.

Teamwork. Debate may facilitate teamwork through creating an environment that is conducive to working collaboratively. The participants recognized and confirmed this merit of debate and acknowledged its influence on their attitude towards it. One participant said:

> It was fun to work together to prepare arguments and rebuttals [during the debate].

Also, a number of participants said that they had learned from each other during debates:

> You learn words from each other. I was impressed by the expression 'up to…' [which a classmate used]."

Debate vs. course book. In the eyes of the participants, debate is an interesting teaching tool that has the ability to pique their interest, arouse their curiosity, and make them experience fun while learning, unlike course books, which are boring and fail to infuse students with

enthusiasm. One participant said that this factor is the most important one that accounts for her liking for debate. She also said:

> I prefer debate over course books because course books lack the ingredients that make debate interesting.

Language proficiency. The participants in general believe in the power of debate in honing their four language skills. One participant said:

> You practice all the skills at the same time.

When asked whether the factor of *language proficiency* was an ingredient that made them like debate, the interviewees answered in the affirmative. One further confirmed this by saying:

> You practice the [four] skills in a fun way [in the debate].

4.5 Conclusion and limitations of the study

The participants' responses in the questionnaire and interviews revealed that the positive attitude that students have towards debate is shaped by the following factors: *active participation*, *challenge*, *teamwork*, *fun*, *critical thinking*, *language proficiency*, and *debate vs. course book*. That is to say, the results revealed that the participants recognized the contribution of all the factors in question to their appreciation of debate.

To make the learning process successful, teachers should invest in the pedagogical tools that students admire and make them experience fun while learning. This study has revealed that students acclaim debate as an interesting teaching tool that energizes them to participate in a rich and engaging learning process. Therefore, debate as a teaching tool should have a place in the pedagogy of L2 teaching.

Given the small sample size of the study, the ability to generalize its findings to all L2 students is restricted. Therefore, further research is needed to confirm these factors and maybe uncover other factors that have not been identified in this study. More research is also needed to further explore the pedagogical benefits of debate and their implications for the pedagogy of L2 teaching.

Chapter 5

Students' Perceived Effect of In-Class Debates in Second Language Learning

This chapter is based upon: 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*(1), 35–57.

Abstract

This study sets out to give voice to students' perceptions and perspectives regarding the effectiveness of in-class debates for developing their English language proficiency across the skill areas of speaking, writing, reading, and listening, as well as vocabulary and grammar. The findings presented are based on questionnaire and interview data collected in a secondary school in the Netherlands from 134 students who debated in 10 sessions. The analysis also drew on debate assessment. The results show that the participants hold a positive attitude towards debate and correlate debate participation with an improvement in all language areas being investigated. Further, analyses reveal that there is a significant correlation between the attitude variable and perceived effects, and between the debate marks and average marks for English. In light of these findings, pedagogical implications and suggestions for future studies are discussed.

Keywords: perceptions, in-class debates, language proficiency, attitude

5.1 Introduction

Debate activities are commonly associated with competitive debate, often in the form of well-established debate leagues and competitions (Aclan & Aziz, 2015a; Jones, 1994). However, debate, as a learning/teaching tool, has recently started to gain ground in education, especially in universities. Many educators use it as a vehicle for accomplishing a range of objectives. Research has attributed a number of benefits to debates. Communication and critical thinking are the two skills commonly identified as the major benefits of debating (Akerman & Neale, 2011; Darby, 2007). Other benefits include developing empathy, evaluating information critically, enhancing disciplinary learning, increasing confidence in public speaking, and fostering research skills (Darby, 2007). Importantly, debate may also hold potential for improving language skills. This study was carried out to investigate second language (L2) students' perceptions of the impact of debating on their language development.

5.2 Features that potentially make debate effective

5.2.1 Interactiveness

Debate is an interactive pedagogical tool. The interaction manifests itself on three levels. Firstly, it takes place on the learner-content level since students interact with information, arguments, and texts. Secondly, the interaction occurs on the learner-instructor level as instructors stimulate debaters and provide feedback on their performance. Finally, the interaction between students demonstrates the learner-learner level. Debates furnish students with fertile ground to cooperate with one another in order to develop a valid justification for their stance and locate weaknesses and flaws in their opponents' arguments.

5.2.2 Competitiveness

Debate immerses students in a challenging clash of arguments (Branham, 1991). This challenge ignites students to perform to the best of their ability. Challenge is one of the features that make students admire debate (El Majidi et al., 2015a, 2015b; Jones, 1994; see Chapter 4). During debates, students push each other to the limit. They challenge each other not only in terms of the strength of the arguments they advance but also in the way they frame them. In other words, their challenge does not stop at the argument strength level, but it extends to the language use level (El Majidi et al., 2015b; see Chapter 4). In the same vein, Mitchell (1998) maintained that "contest round competition is a powerful motivating force that draws in novices and pushes advanced debaters to dizzying heights of professional and academic excellence" (p. 50). Furthermore, Warner and Bruschke (2001) contended that the competitive nature of debate can

even stimulate less enthusiastic students to do research to fare well in debate. In other words, "competition becomes a vehicle for motivating students to research" (Warner & Bruschke, 2001, p. 18).

5.2.3 Active involvement

Involving students in the learning process and captivating their attention is a challenging task that many teachers grapple with. Involvement in the learning process is one of the secrets of a successful learning experience. Dörnyei (2002) argued that task engagement is a key issue in instructed second language acquisition because it is an essential condition for any language processing to occur; that is to say, "if students are not actively involved in the instructional tasks and do not produce a certain amount of language output, L2 learning is unlikely to be effective in developing communicative skills" (Dörnyei, 2002, p. 144). In this connection, Mackey's (1999) study revealed that active participation in interaction yielded progress in terms of L2 production.

Because of this, teachers need to employ viable pedagogical tools that smooth the path for the active involvement of students in the learning process. Darby (2007) stated that debate is one of the tools that facilitates involvement and captivates attention in the classroom, as it involves students in a rich learning process that entails researching, preparing and presenting arguments, evaluating information, and formulating opinions. Likewise, Lustigova (2011) maintained that debate actively involves students in the learning process and that even in its modified and simplified version, debate can be effective for everyone involved. Additionally, Stewart and Pleisch (1998) pointed out that in-class debates motivate the whole class, including passive students.

5.2.4 Attitude

A number of studies have demonstrated that task attitude functions as a filter (Dörnyei, 2002; Dörnyei & Kormos, 2000) and boosts students' motivation to perform to their highest level (MacIntyre et al., 2001; Masgoret & Gardner, 2003). More importantly, recent empirical research has revealed that there is a positive correlation between task attitude and language acquisition (Dewaele et al., 2018; Jansen et al., 2016). Thus, for in-class debates to be effective in second language learning, students should perceive debating as motivational and interesting. In other words, they should have a positive attitude towards debating as a learning tool.

Earlier research has shown that students have a positive attitude towards in-class debates. Stewart and Pleisch (1998) pointed out that debate was consistently credited with a rate above 80% as "the best/most interesting course activity" in their course surveys that lasted four years. They also mentioned that all the students considered debate an indispensable part of the course.

Similarly, Randolph (2007) reported that 85% of her participants recommended debates for future classes and 82% experienced them as fun. Lustigova (2011) remarked 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. She also reported that attendance rose by 20% after the announcement and start of debates. In a similar vein, various other studies reported that the students who participated in debates described them as attractive and instructive (e.g., El Majidi et al., 2015a, 2015b; Gelinck, 2000; Omelicheva & Avdeyeva, 2008; Zare & Othman, 2015; see Chapter 4).

5.3 Debate in the second language context

Several studies have suggested that L2 in-class debates have the pedagogical potential to involve students in a rich learning process (Snider & Schnurer, 2006; Zare & Othman, 2015). For students to fare well in debates and optimize their benefits, good preparation is an indispensable requirement. In the preparation phase, students need to research the topic of debate. By so doing, students not only become knowledgeable about the topic at hand, but they also sharpen their critical reading as well. A proper preparation also entails preparing a case[1] and noting down the arguments the opposing team would likely use. Furthermore, a meaningful debate cannot take place without students critically listening to each other's arguments to spot flaws in them. It goes without saying that speaking skills receive the lion's share of attention in the debate process. In other words, as Zare and Othman (2013) clarified, in-class debates facilitate access to linguistic input and output. This is also confirmed by Lustigova (2011), as she argued that the debates in her study managed to integrate skills "in such a way that they supported and enhanced one another" (p. 26). Additionally, the integrative power of debates saves time, as they enable the teacher to kill up to four birds with one stone (Gelinck, 2000).

Although many studies have applauded the pedagogical qualities of L2 debate, little research has been generated that has demonstrated that participation in debate results in language development (Omelicheva & Avdeyeva, 2008). Littlefield (2001) noted that this dearth of research is particularly noticeable in the secondary school context. It manifests itself in the fact that "very few manuscripts dealing with high school debate have been published in academic journals" (Littlefield, 2001, p. 83). The scarcity of research that Omelicheva and Avdeyeva, and Littlefield pointed out concerns the first language context. In the L2 context, the debate research is scarce in the extreme.

[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.

To establish the effectiveness of L2 in-class debates, further research is needed. Only with research are we able to investigate the full pedagogical potential of this learning/teaching tool. It is true that many studies have claimed that its implications for learning are promising; however, these claims are groundless unless established on empirical grounds. One way of researching the potential effect of in-class debates on learning is through investigating students' perceptions. The need for this research was also called for by Park et al. (2011) who said that "further research is needed on this teaching strategy … also in terms of student perceptions and experiences" (p. 14).

There are two notable studies—Lustigova (2011) and Aclan and Aziz (2015a)—that have explored the effects of in-class debates on L2 students' language development. Lustigova (2011) reported that debate in L2 enhanced her participants' language proficiency. The participants in Aclan and Aziz (2015a) also reported improvement in their language skills after debating. However, Lustigova's findings were based on her observations and those of Aclan and Aziz on what six students reported in an interview. Moreover, both studies were conducted on university students. The studies that have elicited secondary school students' perceptions of the effects of in-class debates on language proficiency are absent. Hence, this study is designed to give voice to secondary L2 students' perceptions and perspectives of the impact of in-class debates on different aspects of language development, namely the skill areas of speaking, writing, reading, and listening, as well as vocabulary and grammar. We opted for this approach, as research has shown that students can have a realistic perception of their own performance and improvement (Ross, 2006; Stefani, 1994). Furthermore, since research has demonstrated that task attitude can function as a driving force variable that can affect students' perceptions and performance (Dewaele et al., 2018; Dörnyei & Kormos, 2000; Jansen et al., 2016), this study also investigated the correlations between attitude towards debate and the perceived effects as well as performance. In short, the study was guided by the following research questions:

RQ.1 What are the secondary school students' perceptions of the effects of L2 in-class debates on their language development?

RQ.2 What is the relationship between attitude towards debate and perceived effects, and between attitude and debate marks? Also, what is the relationship between debate marks and average marks for English?

5.4 Method

5.4.1 Participants

One hundred thirty-four Dutch students (63 males and 71 females) participated in the study. Their ages ranged from 14 to 19. The study spanned a period of two years (four classes participated in the first year and three classes in the second year) and involved seven classes in a Dutch secondary school in Rotterdam. Two groups were lower classes (Grade 9) comprising 38 students, including 19 males and 19 females (aged 14–16). They studied English at the B1 level. The other groups were five upper classes (Grade 11) consisting of 96 students, including 44 males and 52 females (aged 16–19). They studied English at the B1–B2 level. All groups received three English sessions of 50 minutes per week from the first author. For the purpose of this study, one session was devoted to debate, and in the other two sessions, the students received regular lessons based on course books published in the Netherlands. The participants carried out 10 debates of different formats. The group sizes ranged from 16 to 25 students.

5.4.2 Intervention

Debates were staged over the course of 10 weeks and revolved around controversial topics selected by the students. To fare well in debates, students must be well prepared. At least one week prior to debates, the students were informed of the debate topic and the side they would defend. Prior to each debate, the students received one article containing information about the debate topic and were tasked to find and read at least one more article. To ensure that the students prepared sufficiently and properly, they were required to submit their debate preparation form (see Appendix A), which included their cases, arguments for both sides of the topic, and the sources they employed. The students who were not in the debating groups were tasked to make notes of their classmates' performance (see Appendix B).

The study employed debate formats that were previously piloted and refined. We used three debate formats: splitting the class into two teams (one team in favor and the other against), debating in groups of four debaters (two students in favor and two against), and a one-to-one debating format.

5.4.3 Instruments

To measure the effects of in-class debates on language development as perceived by the participants across the skill areas of listening, speaking, reading, and writing, as well as grammar and vocabulary, we developed a questionnaire (Appendix D). Most of the items were modeled on, or adopted from, literature published on language skills development and attitude (e.g., Jones, 1994; Stepp-Greany, 2002; Zare & Othman, 2015).

However, questionnaires do not always conclusively work out why respondents answer in a particular way, nor is it evident whether all respondents interpret items in the same way the researcher intended. To overcome these shortcomings, we conducted in-depth interviews after the study. The interview data were meant to strengthen and enrich the questionnaire data and remedy their weaknesses.

In the present study, 23 semi-structured one-on-one and group interviews were conducted with 65 volunteered students from all the participating classes at the end of the study and were audio-recorded. Eleven interviews were one-on-one and 12 were group interviews (six interviews involved six participants and six involved three participants). The goal was to gain an in-depth understanding of the students' perceptions of, and reflections on, the effects of debates on language proficiency.

During the interviews, the students were asked to comment on the potential effects of debates on their language skills in addition to grammar and vocabulary. When effect on a particular skill was reported, the participants were then asked to elaborate on the scope of this effect and to provide examples. The duration of the interviews ranged from 10 to 27 minutes. Prior to every interview, the students were briefly informed about the objectives of the interview and were assured confidentiality. The interviews were scheduled during school time.

The participants' performance during debate was assessed with a mark by the first author. The assessment took place twice for 81 participants (from all classes). The first assessment took place during the second and third debates and the second one during the last two debates. The assessment covered areas that measure content and oral proficiency aspects, drawing on Brown (2004) and Oros (2007): *constructive speech*, *reasoning*, *rebuttal/clash*, *fluency*, *vocabulary/-variation*, *grammar*, and *preparation*. Each component was assessed on a scale of five points. Because of time constraints, it was difficult to assess each debater's performance twice. It is worth noting that the students were not informed of the moment of the assessment.

5.4.4 Pilot study

Prior to administering the questionnaire to the participants, a pilot study was conducted in accordance with the suggestions and guidelines of Dörnyei and Taguchi (2009) to identify essential revisions and to explore the validity of the scales. A group of students (n = 12), who had already participated in a number of debates, were recruited to complete the draft questionnaire while thinking aloud to check whether the items were interpreted as intended and also to evaluate their clarity, the response format, and the overall appearance of the questionnaire layout.

The questionnaire initially consisted of 35 items representing five scales, with each scale containing five items. The students indicated that a number of items tapped into the same

underlying constructs and that some items evoked a different meaning than the study intended to measure. As a result, a number of items were merged, and all scales were reduced to three items. After implementing all the revisions suggested by the students, nine students were invited again to inspect the revised questionnaire. The students reported that no revisions were necessary. The final version of the questionnaire contained 21 items tapping into seven constructs (attitude, perception of the effects of debate on reading, listening, speaking, and writing skills, and on vocabulary, and grammar) and a few personal questions eliciting background information. The items appeared in random order to avoid annoying the students with repetitive content. Moreover, some items were negatively formulated to force the participants to evaluate each item in its own right.

To establish the reliability of the questionnaire, its final version was piloted with 31 students comparable to the target sample. Item analysis was performed using SPSS to check the internal consistency of each construct. The internal reliability of the constructs exceeded the acceptable threshold (.70). This ensured that the questionnaire did not contain any glitches that could put its reliability at risk in the final administration. The interview was also piloted with three volunteer students to check if it contained flaws, limitations, or ambiguities. After the piloting stage, the necessary revisions and modifications were made.

5.4.5 Procedure

The participants were asked to rate each item of the questionnaire on a five-point Likert scale in terms of its perceived importance from 1 (*strongly disagree*) to 5 (*strongly agree*). To make sure that they filled out the questionnaire appropriately, the initial instructions were read out, the importance of truthful and thoughtful answers was stressed, and confidentiality was promised and assured. The questionnaire was completed a week after the last debate, and its completion took approximately 10 minutes. Subsequently 65 participants were randomly chosen from the seven classes and invited to elaborate further on their perceptions and perspectives of in-class debates.

5.4.6 Data analysis

To answer the first research question, the internal reliability of the scales was first calculated. It was verified by means of Cronbach's alpha coefficient. The internal reliability reached an acceptable threshold on all scales (see Dörnyei & Taguchi, 2009). Cronbach's alphas ranged from .70 to .84. This means that the items of the scales are interrelated and hence measure the same underlying constructs. Then, a one-sample *t*-test was conducted to examine whether the mean scores significantly deviated from the scale center.

To address the second research question, first a paired samples *t*-test was carried out to track any improvement in the assigned debate marks between Assessment 1 and 2. Then, correlational analyses were conducted to investigate the correlation between attitude and perceived effects of debate, and between attitude and debate marks. Further, correlational analysis was also conducted to examine the relationship between debate marks and average English marks. Prior to correlational analyses, preliminary analyses were performed to ensure no violation of the assumptions of normality, linearity, and homoscedasticity.

5.5 Results and discussion

The results are presented and discussed in relation to each research question.

RQ.1 What are the secondary school students' perceptions of the effects of L2 in-class debates on their language development?

As Table 5.1 shows, the mean scores of all constructs exceeded the neutral point (3). They range from 3.63 to 4.32. This indicates that the participants have a positive attitude towards debate and perceptively correlated participating in it with an improvement in their listening, reading, writing, and speaking skills, as well as their vocabulary and grammar. This fact is confirmed by a one-sample *t*-test, which revealed that all mean scores on all scales significantly deviated ($p < .001$) from the scale center. However, as the means clearly show, the participants perceived that the debates had more effects on certain areas of language than others. They perceived that their vocabulary and speaking skills improved the most. On the other hand, listening and grammar were felt to receive the least effect.

Table 5.1 Descriptive statistics for the participants' perceptions of the effects of debates

Constructs	*N*	*M*	*SD*
Attitude	134	4.27	.57
Speaking	134	4.26	.60
Writing	134	3.94	.81
Reading	134	3.88	.77
Listening	134	3.63	.80
Vocabulary	134	4.32	.65
Grammar	134	3.65	.78

We now further discuss the participants' perceptions, drawing mainly on the outcomes of the interviews.

Attitude. The participants in this study have a very positive attitude towards debate ($M = 4.27$). This concurs with previous studies (see Section 5.2.4). In fact, in-class debates possess the

ingredients that make learning an interesting experience. They revolve around issues that students encounter in their daily life, interest them, and require them to take an active role in the learning process, rather than absorbing information passively. In other words, as one participant declared:

> Once you like something [she refers to debates], you learn with pleasure.

The participants also admired debates because they disentangled them from "those boring course books." This finding is not surprising. *Course book vs. debate* emerged as the most important factor that shapes students' positive attitude towards debate in El Majidi et al. (2015a, 2015b; see Chapter 4).

All the interviewees emphasized that debating is an enjoyable experience, though a number of them remarked that it was tough to debate at the beginning because it was frightening to talk in front of the class. However, these students noted that after participating in a number of debates, they managed to gain more confidence and feel less nervous.

Some participants pointed out that they liked debates because they enabled them to improve their language proficiency. These students raise here an interesting point of pedagogical importance. They seem to suggest that perception of task benefit can shape task attitude. In other words, this implies that task attitude correlatively accrues with increasing task benefits. One participant said in this regard:

> I am thankful to debate lessons. Without them I would have [probably] never read English articles … [and because of them] my vocabulary has improved.

No wonder that when students have a favorable attitude towards a particular task, they tend to perform to the best of their ability and mobilize their full effort. This corresponds to the study of Dörnyei and Kormos (2000), who concluded that task attitude functions like a filter and hence can affect students' performance.

All the participants in the interview advocated planning debates on a regular basis but not every week because of the preparation they entail. However, the participants were divided over whether debates should be introduced in other modern foreign languages. The opponents believe that students do not possess enough language abilities to stage decent debates in these languages. On the other hand, the proponents think that even with modest linguistic resources, debates can be interesting.

Speaking. The participants felt that the debates substantially honed their speaking skills (M = 4.26). This did not come as a surprise. Speaking occupies the lion's share of attention in the debate process. It is even practiced in the preparatory phase as debaters try to rehearse their

cases. Previous studies (e.g., Lustigova, 2011; Zare & Othman, 2015) have also reported that L2 debaters improved their speaking as a result of participating in in-class debates.

The interviewees overwhelmingly acknowledged that the debates progressed their speaking proficiency. When asked to indicate how they perceived this progress, some pointed out that they could produce longer constructive speeches towards the end of the debate series than at the beginning as this participant noted:

> During the first debate a lot of students couldn't talk one minute … after five [debate] lessons, everyone was able to do it [talk one minute]. The teacher had to stop us. We were able to talk more than two minutes. At the beginning this was impossible.

It is important to note that our observation corresponds to what the students reported. It was observed that most debaters at the start of debates could hardly produce one minute of constructive speech, while at the end of the series, they effortlessly talked for more than one minute. It was also observed that they managed to frame longer rebuttals in the last debates. These observations hold true for the clashes as well.

Impromptu speaking, or the ability to talk "on the spur of the moment" without time to prepare, was another outcome of debate. A number of students reported that the debates enhanced their ability to think on their feet, and as a result, they banked more on improvising and less on memorizing:

> At the beginning I had to learn [my case] by heart. Later [towards the end of debates] I just improvised. I just wrote keywords, and on the basis of this I improvised my speech. I didn't need to learn by heart anymore as at the beginning.

Another area of speaking that benefited from debating, according to many interviewees, was pronunciation. The interviewees remarked that the debates enabled them to learn the right pronunciation of many words. Lastly, they also noted that the debates endowed their voice with confidence, and they mentioned that the debates enabled them to "take courage in both hands" when talking in front of their classmates.

Writing. The participants also perceived that the debates improved their writing skills with a mean of 3.94. During the debate process, writing was particularly practiced in the preparation phase. Before the actual debates, the debaters were required to write down their cases in their own words and submit them. The students received feedback on their cases and were required to rewrite and improve them. During debates, the students also practiced writing since they were required to make notes of their adversaries' arguments, which they had to process in their rebuttals. It is of paramount importance to note that the perceived effect on writing emanated

not only from the attention writing received in the preparation phase and during debates (making notes) but also from the improvement of other skills. Language skills are interrelated. They promote each other. Carlin and Payne (1995) highlighted the interplay between speaking and writing during the debate process:

> Many of the skills you use to compose good cases are the same as those you need to write well. The practice you gain preparing cases will improve your ability to express your ideas on paper. You will learn how to get your audience's attention, how to organize clearly, and how to support your ideas. (p. 9)

During the interviews, the participants reinforced the perception that the debates developed their writing competence. They reported that the progress they felt was due to the increasing ease with which they were able to write their cases towards the end of the debate series. Some attributed this ease to the increase in their writing fluency as this participant noted:

> At the beginning, I needed much time, up to half an hour, to write down arguments I thought up. Now if I have the arguments, I just need 10 minutes to be done.

Other participants mentioned that they felt the perceived effect on their writing through their ability to compose better letters. Others noticed a decrease in the grammatical mistakes they made and an increase in the ability to compose better coherent texts. It should be noted that almost all the interviewees stressed that the vocabulary they learned in pre- and during-debates also enabled them to write better English.

Reading. The participants recognized the effect of the debates on their reading skills ($M = 3.88$). It is worth reiterating that prior to each debate, the participants received one article from the instructor and were required to find at least one additional article. The students were instructed to read these articles carefully to find the arguments that bolster support for their side as well as the arguments that weaken the "other" side.

Many interviewees felt this effect on their reading skills. This effect manifests itself, according to many of them, in the ability to read more critically and comprehend texts better. One participant summed up this effect succinctly when he said:

> You understand texts better and quickly.

Another participant elaborated on this:

> I understand texts better because before [the debates] I sometimes had to read some paragraphs twice or three times before I could really comprehend them, but now because

> of debates, when I read articles, I comprehend them immediately [without rereading them].

A number of interviewees also maintained that the debates improved their reading speed. They believed that the debates helped them to read more quickly. Moreover, the majority of the interviewees underlined that the debates considerably enriched their vocabulary, which, in turn, enabled them to comprehend texts better.

Listening. The participants also felt that the debates developed their listening skills ($M = 3.63$). However, this skill was accorded less effect than the other three skills. Unlike other skills, listening is practiced almost only during debates. During debates, the participants had to listen carefully and critically to their opponents so as to be able to locate flaws in their reasoning. These flaws were needed to compose strong rebuttals and conduct powerful clashes. Consequently, this pushed the participants to summon up their concentration and energy in order not to miss any weaknesses in their opponents' arguments that could tip the scales in their favor.

The majority of the interviewees indicated that the debates honed their listening skills. Some accounted for this effect by means of comparing their listening ability at the start of the debate series and at the end, like this interviewee:

> If I compare the first and the last [debate] lesson, I can see that in the first debate, I almost didn't understand anything, while at the end, I could understand almost everything.

Other interviewees reported that the debates eased their understanding of English language TV programs, like this interviewee who praised the debates for enabling him to understand films in English better:

> I have noticed that when I watched films in the past, I used to read subtitles. I understood what it was about, but I had to read for certainty. But now when I watch, I understand better. I listen better now. I understand what people say … during debate lessons I had to listen with concentration.

Vocabulary. The participants evaluated the impact of the debates on their vocabulary acquisition with a high mean ($M = 4.32$). With this mean, the participants clearly signaled that the debates substantially enlarged their vocabulary. It is common knowledge that vocabulary forms the backbone of the four language skills. Debate proves to be fertile ground for the acquisition of new vocabulary and its active implementation. Prior to each debate, the participants were required to read at least two articles. The participants were instructed to read them carefully to

extract arguments. To meet this end, it was essential to understand, if not all, at least the majority of vocabulary. Such a process leads to decoding the unknown words in a context that bears importance for the students. Additionally, the debate environment facilitates the implementation of the newly learned words, thereby consolidating their acquisition.

During the interview, the participants massively acknowledged and highlighted the role the debate process played in extending their vocabulary:

> Each time I read an article, I learn seven to eight words I didn't know. I searched them and wrote them down. This [process] allowed me to know these words better, which I used during rebuttals. And during debates, I had to write down those [new] words of my classmates.

Most interviewees pointed out that they learned five words on average from each debate. They also pointed out that they learned words from their classmates during debates.

Grammar. The participants also felt that the debates improved their grammar. However, the perceived effect ($M = 3.65$) was comparatively not strong. Grammar received attention mainly in the post-debate stage. At the end of each debate, the participants submitted their cases, on which they received feedback from the instructor. During debates, the participants were also instructed to note down the mistakes of their classmates and improve them.

During the interviews, the participants reported that they felt that the debates improved their grammar more than they had revealed in the questionnaire. The majority of the interviewees associated improvement in their grammar with the feedback they received from the instructor on their cases as this interviewee declared:

> I have noticed that I don't make the mistakes I made before anymore and were improved [by the instructor] and which I had to rewrite. If I make them, then it's because of being sloppy, but my grammar has improved. I talk better [English] now. My sentence construction has also improved.

Our observation also corroborates the perceived effect of corrective feedback on the participants' grammatical competence. The cases that were submitted towards the end of the debate series contained fewer errors than the ones being handed in at the beginning. The participants' revelations and our observations provide further evidence to the effectiveness of corrective feedback and also cast light on another potential of debate, namely improving students' accuracy.

RQ.2 What is the relationship between attitude towards debate and perceived effects, and between attitude and debate marks? Also, what is the relationship between debate marks and average marks for English?

With regard to debate assessment, the majority of the students received a positive assessment/mark. The marks ranged from 5/10 to 9.2/10. As for the students whose performance was assessed twice ($n = 81$), a paired samples *t*-test was conducted to compare the first and second assessment. Statistically, as Table 5.2 demonstrates, all the debate and performance aspects being assessed significantly improved from Assessment 1 to Assessment 2.

Table 5.2 Paired samples *t*-test between Assessment 1 and Assessment 2

Components	Assessment 1	Assessment 2	*t*	*df*	*P* (two-tailed)
Constructive speech	3.01 (.63)	3.67 (.54)	-11.46	80	< .001
Reasoning	3.04 (.59)	3.61 (.51)	-4.29	80	< .001
Rebuttal/clash	3.07 (.68)	3.58 (.58)	-8.23	80	< .001
Fluency	3.05 (.59)	3.56 (.59)	-9.67	80	< .001
Vocabulary/variation	3.01 (.60)	3.49 (.56)	-8.98	80	< .001
Grammar	3.11 (.66)	3.60 (.64)	-8.47	80	< .001
Preparation	3.40 (.78)	3.62 (.71)	-4.27	80	< .001

The relationship between attitude towards debate and perceived effects was investigated using correlational analysis. There was a significant, positive correlation (at $p < .001$ level) between attitude and perceived effects in all skills (listening, $r = .47$; reading, $r = .41$; speaking, $r = .38$; writing, $r = .32$) as well as vocabulary ($r = .47$) and grammar ($r = .38$). These correlations seem to suggest that the attitude variable tends to positively shape the students' perceptions of the effects of debate on their language development. Thus, the more positive the attitude towards debate is, the higher the perceived effect is. This is not a surprising finding since students tend to put strenuous effort into tasks they view as positive. This filtering effect can have a positive impact on fluency, accuracy, and complexity of the output (Dörnyei & Kormos, 2000). This finding is also in agreement with Dörnyei (2002), whose study revealed that students with a high task attitude outperformed those who had a low task attitude. Because of this, Schulz (1996) emphasized the importance of attending to students' attitude towards teaching practices since this has a weighty bearing on the effectiveness of teaching. In other words:

> to establish pedagogical credibility and increase their students' commitment to and involvement in learning, teachers make an effort to explore students' beliefs about language learning and to establish a fit between their own and their students' expectations. (Schulz, 1996, p. 343)

The relationship between attitude and debate marks was also of interest. The correlation between these two variables was not significant ($r = –.06$, $p < .49$). This indicates that students of all abilities, low or high, appreciate debate. This is an interesting result because not only do the higher-ability students seem to favor debate but apparently the lower-ability students do as well. This is consistent with previous research (e.g., Stewart & Pleisch, 1998) that demonstrated that the lower-ability and passive students also admire debate.

Lastly, the relationship between debate marks and average marks for English was also investigated using correlational analysis. There was a strong, positive correlation between the two variables ($r = .58$, $p < .001$), with high debate scores leading to higher marks for English. This finding indicates that debate marks can largely predict students' marks for English. The strong correlation also makes a case for the reliability of debate marks in reflecting the debaters' competence and improvement in English. It is worth mentioning that the average mark for English was a mark built on a number of marks, ranging from seven to sixteen marks, for the four skill tests in addition to grammar and vocabulary tests.

5.6 Conclusion and limitations

In this study we have investigated students' perceptions of the impact of debates on their language development, and the relationship between attitude towards debate and the perceived effects as well as assessed performance. The results demonstrate that debate is highly appreciated by students and that they perceptively correlate participating in it with an improvement in language proficiency. This finding is perhaps not surprising, given that similar results have emerged from other studies. However, unlike previous studies whose participants were recruited from universities, our participants were secondary school students. Moreover, in this study, we have managed to obtain a larger data sample and to draw on different data sources (questionnaire, interviews, observations, and marks), a fact that renders the ensuing findings more reliable and thus solidly champions the facilitative role of debates in L2 learners' language development. Perhaps this study distinguishes itself more from previous studies (on students' perceptions of the effects of L2 debate on language proficiency) by tracking the effects under study in correlations between various variables, including attitude, perceived effects, and assessed performance. Including marks in the correlational equations has provided more insights into the scope of the perceived effects.

The current study has not only proved that in-class debates facilitate the integration of skills, but it has also substantiated that they have the potential to improve them. The questionnaire and interview data point in the direction that in-class debates offer opportunities to involve students in an enjoyable and rich learning experience. In this learning experience, the

four language skills in addition to vocabulary and grammar are called into play in a way that strengthens and enhances each other. During the interviews, the interviewees repeatedly reinforced this and elaborated further on how the skills interrelated. This is an important finding since various studies have contended that tasks whose completion involves input and output enhance language acquisition. Pedagogically, this finding underscores the utility of employing tasks that lend themselves to skill integration. This not only saves time, but it enables skills to consolidate each other as well.

It is important to note that the interviewees revealed that the language development they had felt during debates extended to other learning situations. They reported, for example, that the debates helped them to speak more fluently outside the classroom, score better on speaking tests, compose better letters in terms of style and grammar, improve their reading pace, gain better results in reading comprehension tests, and understand films better. Additionally, they highlighted that they recognized many of the words they had learned from the debates in different contexts and also used them. Indeed, these revelations add another dimension to the perceived effects, namely they make a strong case for the transferability of the performance gains to a wider range of contexts. They also provide a case for vocabulary retention. In short, these facts seem to bear witness to the external reliability of the emerging results.

It should be noted that we believe that the transferability of the gained learning to "new" contexts is also facilitated by the positive attitude that the students have towards debate. In other words, we take the view that once students perform a task they "enjoy," learning transferability and retention have a larger probability of success. Therefore, we believe that learning transferability and retention may correlate with task attitude.

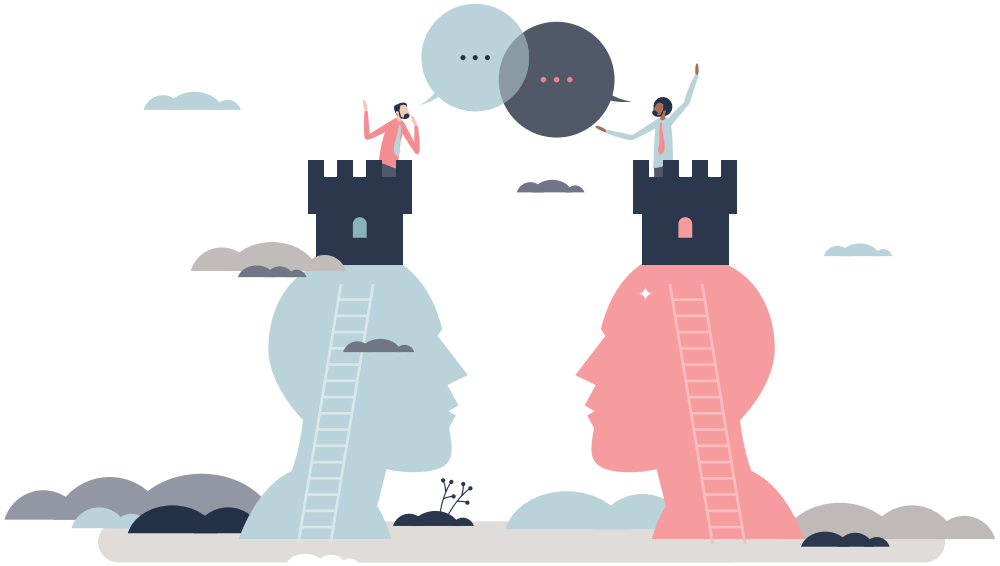
It is imperative to take heed of students' voices and perceptions of what constitutes effective learning. We should not underestimate their ability to judge and reflect on the quality and effectiveness of learning tasks. Their views and opinions should be treated as an important piece of information that needs to be taken into consideration in the pedagogical decision-making process. Consideration should also be given to students' task attitude. A positive task attitude can enhance students' motivation and can lead to L2 acquisition. A positive task attitude can also create a relaxing atmosphere in which learning becomes a pleasurable experience.

Admittedly, this study has limitations. Though the study involves seven classes, it was conducted in merely one secondary school. Moreover, debate performance was rated solely by the instructor. Another relevant limitation is that though some studies remarked that students' perceptions of their language improvement provide fairly reliable information, empirical and longitudinal research is still needed to accurately gauge the perceived effects. Nevertheless, the results have been consistently confirmed through different data sources, a fact that makes them powerful enough to suggest some valid tendencies and patterns. The study therefore provides

insights for schools and universities currently employing or contemplating the implementation of in-class debates. The study also attempted to bring to light a range of pedagogical qualities of debates that make a strong case for including them in L2 curricula. To the best of our knowledge, this is the first study that has elicited L2 students' perceptions of the effects of debate on language development in the secondary school context. Thus, more research is needed to further investigate these perceptions in other educational contexts. Such studies may contribute to a knowledge base that will help teachers to improve students' L2 learning by debate and also to deepen our insight into effective L2 pedagogy.

Chapter 6

Debate as a Pedagogical Tool for Developing Speaking Skills in Second Language Education



This chapter is based upon: El Majidi, A., de Graaff, R., & Janssen, D. (in press). Debate as a pedagogical tool for developing speaking skills in second language education. *Language Teaching Research*. https://doi.org/10.1177/13621688211050619

Abstract

Many secondary school students' L2 speaking skills suffer from deficiencies. The effects thereof are detrimental to their academic and career opportunities in a globalized world that highlights the importance of oral communication skills. Debate has been considered a potentially effective speaking pedagogical tool that can scaffold learning processes in ways that can lead to language development. This study investigates the effects of a debate intervention on English L2 speaking competence of Dutch secondary school students. Following a pretest–posttest control group design, we elicited speech samples from opinion tasks that we coded in terms of measures of speech quantity, fluency, complexity, accuracy, and cohesion. Multilevel analysis results indicate that after the intervention, the intervention group produced more language that was more fluent, accurate, coherent, and lexically more sophisticated relative to the control group. These findings, which have significant implications for L2 speaking development, are discussed in relation to specific characteristics of L2 debate pedagogy.

Keywords: in-class debate, L2 speaking skills, speaking development, secondary education, L2 speaking instruction

6.1 Introduction

The ability to speak a second language (L2) properly is an arduous task considering the interwoven factors that come into play when acquiring this ability (Richards & Renandya, 2002; Romaña Correa, 2015; Shumin, 2002). Speaking is a cognitively and socially taxing skill; it entails encoding and expressing thoughts in speech streams that make sense and are contextually appropriate (Goh, 2017). The need to conceptualize, formulate, and articulate thoughts (see Levelt, 1989) demands a lot of cognitive space in the working memory, a fact that hinders learners from adequately attending to all aspects of speech (Skehan, 1998).

Richards and Renandya (2002) stated that "a large percentage of the world's language learners study English in order to develop proficiency in speaking" (p. 201). Learners need to speak English well for better academic and career opportunities in a globalized world that highlights the importance of oral communication skills. Nonetheless, despite its importance, L2 speaking practice has been marginalized in many educational contexts (Baker & Westrup, 2003; Goh, 2017; Thornbury, 2005). As a result, it is no wonder that many studies (e.g., Zare & Othman, 2015) have expressed concern about the speaking ability of L2 learners.

L2 speaking in Dutch secondary schools faces similar neglect (Haijma, 2013; West & Verspoor, 2016). Many students believe that speaking is insufficiently trained in class (Fasoglio & Tuin, 2017) and feel frustrated that they are not able to express themselves fully in the target language after many years of instruction (Haijma, 2013; Piggott, 2019). As a result, many secondary school students' speaking skills suffer from deficiencies, and the effects thereof are more noticeable and detrimental at the university and college stage (Beeker, 2012).

Dutch teachers and students alike have expressed their concern about the current situation. Interviews with Dutch secondary school teachers of English revealed that speaking skills receive the least attention in comparison to other skills in their teaching practice. The teachers ascribed the negligence to the absence of viable teaching tools that would enable their students to effectively practice speaking (see also Fasoglio & Tuin, 2017). In addition, Brown (2009) has maintained that in the L2 context, instructors are confronted with the challenge of finding ways that ensure language development within limited time and budgetary constraints. To overcome these limitations, Brown (2009) proposed employing innovative instructional tools like debates, as their gains "can equal if not exceed uptake that occurs in extended immersion environments" (p. 547).

Debate has been considered a potentially effective pedagogical tool for speaking, which can scaffold and feed the learning process in ways that can lead to language development (e.g., Lustigova, 2011; Stewart, 2003). Speaking occupies the lion's share of attention during debate. In addition to planned speech, debates involve a lot of impromptu speaking, as debaters have

to think quickly and respond to opponents' arguments, especially during the "clash" stage (see intervention section).

Various studies have reported improvements in students' speaking competence after participation in debates. In El Majidi et al.'s study (2018; see Chapter 5), the debaters perceived that debate improved their speaking skills with a mean of 4.26 on a five-point Likert scale. All the respondents in the study of O'Mahoney (2015) found that debates honed their speaking skills. Also, the participants in Zare and Othman's (2015) and Al-Mahrooqi and Tabakow's (2015) studies reported improvements in their speaking skills as a result of debating. Nevertheless, all the studies that have correlated debate participation with oral competency development were based on self-reported data, questionnaires, and interviews, in addition to instructors' observations. Experimental evidence that substantiates the existing anecdotal data is notably absent. The main objective of this study is, therefore, to provide some empirical evidence about the effects of an in-class debate intervention on various aspects of speaking competence, employing a pretest–posttest design. Without empirical evidence, the claimed effects of debate on speaking skills remain groundless. This study is premised on the hypothesis that debate pedagogy constitutes an effective avenue for enhancing many areas of L2 speaking skills. We discuss the rationale behind this hypothesis in the next section.

6.2 Debate as an effective L2 speaking pedagogy

6.2.1 Theoretical grounds

A number of theoretical approaches to L2 acquisition provide a rationale for assuming a potential effect of debates on speaking skills. For example, the Interaction Hypothesis of L2 acquisition (Long, 1996, 2018) pointed out that interactive tasks, such as conversations, set the stage for the negotiation of meaning and that through this channel input and output are connected in a productive way. Likewise, Ellis and Shintani (2014) maintained that interaction can operate "as a source of input and opportunities for output which foster the internal processing that results in acquisition" (p. 194). Interaction provides learners with multiple opportunities to negotiate meaning and form in ways that lead to L2 development (Loewen & Sato, 2018). Gass and Mackey (2015) concluded that "there is a robust connection between interaction and learning" (p. 181).

Debating involves meaningful multilevel interactions (i.e., interaction with content, learner-learner, and learner-instructor interactions). These interactions, which are fueled and enriched by the competitive atmosphere of debates, help learners to notice language gaps and accordingly modify and refine their L2 output. Wade (1998) eulogized the efficacy of the debate

induced interactions, stating that "there are certainly trends in education which encourage interactive and dialogic pedagogies, but few are as potent as debate" (p. 63).

The Output Hypothesis of Swain and Lapkin (1995) provides another theoretical perspective that supports that debates could foster L2 acquisition. They argued that output provides learners with unique opportunities to process language. Output can assist language learning through prompting learners to notice their language gaps, testing out hypotheses (i.e., using forms that are at the cutting edge of their linguistic ability), and reflecting consciously on forms. Engagement in bidirectional output (as is the case in debates) highlights gaps in L2 learners' interlanguage system and hence facilitates attending to the problematic areas in their language (Swain, 2013; Swain & Lapkin, 1995). Benati (2017) argued that involving L2 learners in structured collaborative output tasks can "facilitate the accurate and appropriate use of language forms and structures" (p. 389).

Debating, by its nature, prompts a great deal of oral output as debaters challenge each other's perspectives and feel the urge to outshine each other's arguments and how they frame them (El Majidi et al., 2018; see Chapter 5). In addition to oral output, debate pedagogy may also induce a considerable amount of written output that can boost the oral output (i.e., speaking skills). Furthermore, the debate environment not only raises consciousness about linguistic deficiency, but it also stimulates experimenting with new forms as well as using language consciously (El Majidi et al., 2020; see Chapter 7).

6.2.2 Pedagogical grounds

Speaking activities are commonly considered as "PRACTICE activities rather than LEARNING activities" (Goh, 2017, p. 250). Manchón (2011) hypothesized that the act of writing holds a language learning potential, given that "composition writing elicits attention to form-meaning relations that may prompt learners to refine their linguistic expression—and hence their control over their linguistic knowledge" (Cumming, 1990, p. 483). We make a similar assumption about speaking in our debate intervention, as it fits this pedagogical mold. We believe that the act of speaking holds a comparable learning potential.

Debates create a genuine environment for a meaningful, functional, and purposeful use of the target language. In debates, students argue with a communicative and functional purpose in mind: defending their proposition and weakening that of their opponents. Attaining this goal necessitates the use of accurate and sophisticated language. As we shall see, in our debate intervention the act of speaking is not an end in itself, but it functions as a vehicle for synthesizing and analyzing arguments and as "a task through which language practice can be orchestrated" (Stewart, 2003, p. 15). Anderson (2016) stated that "it is hard to imagine a more harmonious integration of content and language skills than in the teaching of debate" (p. 76).

In short, the debate environment promotes the interface and synergy of two speaking perspectives: learning-to-speak and speaking-to-learn perspectives.

Debate activities place students at the center of learning, with the teacher assuming the role of a coach, advisor, and facilitator. Blumberg (2009) argued that when students are central in the learning process, they are empowered to gain benefits, such as higher rates of content retention, interaction, enjoyment of class activities, and a deeper understanding of material. Similarly, Emaliana (2017) pointed out that student-centered learning provides opportunities for a "conducive atmosphere of learning, dynamic classroom activities, and [offers opportunities] to do autonomous learning" (p. 63). Debates lend themselves readily to the philosophy of student-centered pedagogy. A well-designed debate pedagogy grants students tools and power to manage the learning process with minimal interventions on the part of the instructor. In debates, students do most of the talking and thinking, which promotes deep learning (Bellon, 2000; Cinganotto, 2019).

In addition, debates promote a healthy competitive atmosphere that serves language learning in many ways (Cinganotto, 2019; Warner & Bruschke, 2001). The inherent competitive atmosphere of debate fuels students to generate rich and lengthy negotiations. What is more, research has shown that students hold a positive attitude towards debates and describe them as fun and instructive (El Majidi et al., 2015b; see Chapter 4). This positive task attitude is beneficial to learning, as empirical research has revealed that there is a positive correlation between task attitude and language acquisition (Dewaele et al., 2018).

6.3 This study

Taken together, several theoretical and pedagogical perspectives on L2 acquisition indicate that debate can be a fruitful avenue for oral language learning. Yet, only a limited body of research has investigated how debates can affect oral competence (Omelicheva & Avdeyeva, 2008). Littlefield (2001) noted that this dearth of research is particularly noticeable in the secondary school context. It manifests itself in the fact that "very few manuscripts dealing with high school debate have been published in academic journals" (Littlefield, 2001, p. 83). The paucity of research that Omelicheva and Avdeyeva, and Littlefield have pointed out concerns the first language (L1) context. In the L2 context, the debate research is scarce in the extreme (Al-Mahrooqi & Tabakow, 2015). No empirical study, to the best of our knowledge, has examined the impact of L2 debate instruction on oral proficiency across all main dimensions of speech production, including speech quantity, fluency, accuracy, complexity, and cohesion dimensions in well-controlled empirical designs. Providing empirical based evidence may stimulate instructors to

employ debates in their teaching practice. The current study was guided by the following research question:

RQ.1 What are the effects of debate pedagogy on different aspects of L2 speaking proficiency, including speech quantity (i.e., number of words and speaking time), fluency, syntactic complexity, lexical complexity, accuracy, and cohesion in secondary education?

In light of the extant literature on debate as well as the theoretical and pedagogical grounds discussed above, we hypothesized that debate pedagogy would have a positive effect on the assessed aspects of speaking proficiency.

To test our hypothesis, we conducted an intervention with a pretest–posttest control group quasi-experimental design. The source of data was speech samples elicited through opinion tasks produced by the students on two occasions: at the beginning of the intervention (pretest) and towards the end of the intervention (posttest).

6.4 Method

6.4.1 Setting and participants

This study reports on part of a larger study, also reported in chapters 7 and 8. The study was conducted in eight intact classes in three public secondary schools in the Netherlands (N = 147), located in three urban areas. Five classes were in the fifth year of higher general secondary education track (or *havo 5* in Dutch[1]) (n = 89), and three classes were in the fourth year of preuniversity secondary education track (or *vwo 4* in Dutch[1]) (n = 58). Five classes served as the intervention group (n = 96) and three as the control group (n = 51). There were 88 female and 59 male students aged between 15 and 18. To ensure the comparability of the groups, both the intervention and control groups consisted of both havo 5 and vwo 4 classes.

The English proficiency level (including speaking) of all classes (including havo 5 and vwo *4* classes) spanned on average B1 (the intermediate level in the Common European Framework of Reference level (CEFR)) as estimated by their teachers. With the exception of one intervention class that received on average two English sessions of 50 minutes per week, the other classes received three sessions of 50 minutes. Both groups received regular instruction consisting of activities dealing with the four language skills. For the purpose of this study, the intervention students participated in one debate a week. During that session, the control students received extra regular instruction in which the four language skills were further practiced.

[1] For more information about the Dutch education system, see: https://www.nuffic.nl/en/subjects/education-in-the-netherlands/#secondary-education. See also Michel et al. (2021).

6.4.2 Intervention group

Our debate task design was validated in a previous study following the principles and guidelines of educational design research. Students in the intervention group participated in 10 debates, one per week. The topics of debates (e.g., the right to bear arms) were selected in consultation with the debating students, who received one week of preparation time for each debate.

Each debate consisted of three stages: pre-debate, during-debate and post-debate. In the pre-debate stage (prior to each debate), the students received a reading assignment (article) relevant to the topic under debate and were asked to find and read one additional article. We instructed the students to summarize the articles and to write a case[2] in which they had to defend their standpoints (see Appendix B).

During actual debates, each student presented a speech and a rebuttal (in which the arguments of opponents were addressed) and participated in a clash. While listening to each other, the students were instructed to note down mistakes and the words they learned from each other's contributions. 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: speech, rebuttal, and clash (See Chapter 3.4.2). In the post-debate stage, the teachers provided feedback on the students' written cases and asked them to revise and resubmit them.

6.4.3 Control group

For the purpose of this study, while the intervention students were involved in the debate intervention (once a week), the control students received lessons based on course books (during that session) in which the four language skills were further practiced. These course books, which were published in the Netherlands, were the same for the intervention and control groups. In each of these course book lessons, one language skill is mainly targeted. However, even when speaking is not the main focus of the lesson, it often figures in many lessons, for example, to activate prior knowledge. As to speaking lessons, the control group mainly practiced traditional controlled face-to-face discussions,[3] in addition to discussing newspaper articles they read weekly.

[2] 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.

[3] An example of these tasks is as follows: Study the useful phrases about having a personal conversation. Have a conversation about your holiday nightmare. Your classmate tries to comfort and help you. Switch roles when you're done.

6.4.4 Procedure

We used so-called "opinion tasks" involving different topics to elicit the students' oral performance in the pretest and posttest. Opinion tasks (in which students have to argue in favor of/against a point of view) induce learners to focus on meaning and are accessible to middle schoolers (Dobbs, 2014). In addition, argumentative tasks are "flexible in terms of content as speakers could conceptualize their own arguments relating to the topic" (Suzuki & Kormos, 2020, p. 161). Furthermore, opinion tasks lend themselves more readily to the elicitation and assessment of cohesion than, for example, narrative tasks.

The opinion tasks in the current study consisted of different accessible topics (e.g., smoking should be banned) (see Malloy et al., 2020). These topics were randomly assigned to students. This means that all students from both groups had the same chances of choosing these topics. We made sure that the students did not receive the same prompt during pre- and posttest. The task prompts and conditions had been previously piloted with similar students and proved to be suitable for our participants. The students were explicitly instructed to provide as many arguments as possible. Before performing the task, seven minutes were allotted to planning, and there was no time limit during the performance. Parental permission forms were obtained prior to the beginning of the intervention.

6.5 Data analysis

To assess the quality of the participants' speech samples in the pretest and posttest, we used a variety of measures tapping into different aspects of performance. In the field of second language acquisition, L2 speaking skills have been conceptualized as a composite construct spanning various linguistic areas, including accuracy, fluency, and complexity (Skehan, 2009), as well as cohesion (Hyland, 2005; Lee & Subtirelu, 2015). For this reason, we analyzed the speech samples produced for indicators of fluency, syntactic and lexical complexity, accuracy, and cohesion. In addition, we also took the quantity of performance (the amount of oral output) into account because research has shown that the amount of speech produced by learners can partly reflect their language ability (see, for example, Li et al., 2015). The measures were a mixture of automatically coded features and measures that required hand coding.

In line with procedures used in previous studies (e.g., Derwing & Munro, 2013; Derwing, Munro, et al., 2014), speech samples should last at least 20 seconds for consideration for fluency analysis, after initial hesitations have been removed. To be considered for the analysis of the other measures, we selected a cut-off of 50 words, as speeches of fewer than 50 words do not provide sufficient linguistic information to reliably assess the student's language aspects relevant to our study (e.g., Crossley, McNamara, et al., 2015).

The speech samples were transcribed verbatim using PRAAT (Boersma & Weenink, 2016) by the first author and checked three times at intervals of approximately one month. Nonverbal fillers such as "eh" and "um" were transcribed and treated as filled pauses. A pause was defined as silence or a nonverbal filler of 250 ms or longer (De Jong, 2016). After pruning the students' transcripts by excluding filled pauses, verbatim repetitions, false starts, and self-corrections, we segmented them into AS-units (analysis of speech) following the guidelines of Foster et al. (2000). An AS-unit is defined as "a single speaker's utterance consisting of an independent clause, or subclausal unit, together with any subordinate clause(s) associated with either" (Foster et al., 2000, p. 365). All units of analysis and errors were manually identified by the first author.

To check the interrater reliability for the hand-coded measures, a research assistant, who was masked to condition, verified a randomly selected 25% of the data after training and discussion with the first author. Cohen's kappa indices for intercoder reliability were high for the assessed measures: .94 for mistakes identification, .98 for mistakes categorization and discourse markers, and .97 for error-free clause identification.

6.5.1 Measures

The speech samples were coded for speech quantity, fluency, syntactic and lexical complexity, accuracy, and cohesion in the following ways.

Speech quantity. Following Freed et al. (2004), Li et al. (2015), and Lys (2013), we employed two indices to measure the quantity of oral production: production time (measured in seconds), which refers to the duration of the student's speech, and the total number of words produced in each speech sample.

Fluency. Tavakoli and Skehan (2005) suggested that three dimensions of fluency—speed, breakdown, and repair—best capture the characteristics of different aspects of temporal fluency. Following De Jong (2016), we used the following indices, which we measured via PRAAT software (Boersma & Weenink, 2016):

- Speed fluency was operationalized as inverse articulation rate (i.e., mean duration of syllables) (speech time [excluding pauses]/the number of syllables).
- As to breakdown fluency, we employed two indices: mean length of pause (number of silent pauses/speech time) and the number of filled pauses (number of filled pauses /speech time).

- For repair fluency, we employed two measures: number of repetitions (number of repetitions/speech time) and number of repairs (number of repairs and restarts/speech time).

Syntactic complexity. Norris and Ortega (2009) have argued that syntactic complexity is multifaceted and therefore should be measured multidimensionally. To capture the multidimensionality of syntactic complexity, we have picked out three measures that tap in three dimensions that were frequently assessed in previous studies:

- Global complexity (i.e., number of words per AS-unit)
- Subordination, (i.e., mean number of clauses per AS-unit)
- Length (i.e., mean length of clauses)

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 because it is less affected by text length and "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, p. 69). Word length is widely utilized 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 word frequency index, which calculates the mean logarithmic frequency for all words, describes how often particular words occur in the English language, drawing on the CELEX database (Baayen et al., 1995). A lower word frequency thus indicates higher sophistication.

Accuracy. To measure accuracy, we have chosen the ratio of error-free clauses, which is widely recognized as a reliable global measure for tracking changes in accuracy (Skehan & Foster, 1999; Tavakoli & Skehan, 2005) and which is also considered suitable for an experimental design (Skehan & Foster, 1999). In addition, we computed ratios for four local accuracy measures tapping into different linguistic categories (see Yoon & Polio, 2017 for the operationalization and examples of the first three measures)[4]:

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

- Error-free clauses (EFC)
- Syntactic errors per 100 words
- Morphological errors per 100 words
- Prepositional errors per 100 words
- Lexical errors per 100 words

Cohesion. The present study also casts light on discourse features of speaking performance through investigating the coherence of the produced speech samples. Since the coherence of discourse is enhanced by markers of cohesion (Halliday & Hasan, 1976; Suzuki & Kormos, 2020), which are important in speech, as they help interlocutors to interpret the conveyed discourse (Tanskanen, 2006), we decided to investigate the use of a number of cohesive devices, which are relevant to argumentative texts, adopting Hyland's (2005) framework on interactive metadiscourse. Research has revealed that the frequency and diversity of the metadiscourse markers below significantly reflect the quality of argumentative texts (Qin & Uccelli, 2016; Uccelli et al., 2013). Following Hyland's procedures, we coded speeches for four types of organizational markers, in addition to their diversity of type (cf. Dobbs, 2014; Qin & Uccelli, 2016; Uccelli et al., 2013) and token (El Majidi et al., 2020; see Chapter 7):

- Frame markers: markers that mark the sequence of arguments or counterarguments (e.g., firstly, secondly)
- Code glosses markers: markers that introduce an example or paraphrase (e.g., for instance, in other words)
- Transition markers: markers that mark additive, adversative, or causal relations between clauses (e.g., besides, 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 conclusions (e.g., to sum up, all in all)
- Markers diversity token: diversity of markers in terms of token
- Markers diversity type: diversity of markers in terms of type

6.5.2 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). 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 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. This means that we need to focus on the interaction of time (pretest vs. posttest) × group (intervention vs. control group). For these reasons, we limit ourselves to reporting interactions.

6.6 Results

To test our hypothesis, we obtained two scores (pretest and posttest) for each learner's speech sample produced in each condition (intervention and control groups) and for each measure. Table 6.1 presents the descriptive statistics (means and standard errors) for each measure at pretest and posttest. What can be noticed at first sight is that the means improved on most measures at the posttest in the predicted direction.

Table 6.1 Means and standard errors across time and condition

Measures	Index	Intervention (n = 96)		Control (n = 51)	
		Pretest	Posttest	Pretest	Posttest
Quantity measures	Speech quantity	25.97 (3.08)	55.43 (3.11)	32.27 (4.06)	35.69 (4.07)
	Number of words	59.22 (7.02)	128.44 (7.09)	67.88 (9.25)	75.17 (9.25)
Fluency	Inverse articulation rate	0.224 (.00)	0.217 (.00)	0.230 (.00)	0.232 (.00)
	Number of filled pauses	0.21 (.02)	0.18 (.02)	0.23 (.03)	0.23 (.03)
	Mean length of pauses	0.63 (.02)	0.60 (.01)	0.61 (.02)	0.59 (.02)
	Number of repetitions	0.06 (.01)	0.06 (.01)	0.05 (.01)	0.06 (.01)
	Number of repairs	0.06 (.01)	0.06 (.01)	0.06 (.01)	0.04 (.01)
Syntactic complexity	Global complexity	13.83 (.60)	13.40 (.47)	12.80 (.69)	13.98 (.71)
	Subordination	1.22 (.16)	1.19 (.14)	1.09 (.19)	1.20 (.20)
	Length	6.31 (.17)	6.54 (.15)	6.19 (.21)	6.49 (.21)
Lexical Complexity	MTLD	52.44 (2.56)	60.18 (1.93)	55.70 (2.92)	55.96 (3.04)
	Word frequency	3.20 (.02)	3.12 (.02)	3.18 (.02)	3.18 (.02)
	Word length	4.13 (.04)	4.27 (.03)	4.21 (.05)	4.32 (.05)
Accuracy	Error-free clauses	0.70 (.02)	0.79 (.02)	0.76 (.03)	0.75 (.03)
	Lexical errors	0.92 (.17)	0.51 (.13)	0.65 (.19)	0.88 (.19)
	Syntactic errors	0.94 (.14)	0.48 (.10)	0.85 (.16)	0.56 (.16)
	Morphological errors	3.00 (.29)	2.42 (.22)	2.48 (.33)	2.39 (.34)
	Preposition errors	0.46 (.12)	0.25 (.09)	0.57 (.13)	0.41 (.14)
Cohesion	Transition markers	3.58 (.39)	4.56 (.32)	2.75 (.45)	3.85 (.47)
	Frame markers	0.20 (.14)	1.24 (.11)	0.25 (.16)	0.31 (.16)
	Gloss markers	0.16 (.11)	0.54 (.08)	0.22 (.12)	0.24 (.13)
	Conclusion markers	0 (.04)	0.17 (.03)	0 (.04)	0 (.04)
	Diversity type	1.29 (.09)	2.28 (.07)	1.22 (.10)	1.36 (.11)
	Diversity token	2.71 (.21)	4.61 (.15)	2.33 (.24)	2.70 (.25)

To test whether there were statistically significant differences between the two groups' performance after the intervention, we conducted MLM analyses. Where statistically significant results were achieved ($p < .05$), Cohen's *d* effect sizes (ES) are provided. The MLM results are presented in relation to each measure in Table 6.2.

Table 6.2 Multilevel analysis results

Measures	Index	Fixed effects				Random effects	
		Denominator df	*F*	*P* (one-tailed)	*d*	Variance within class	Variance between classes
Speech quantity	Q. Speech	277.886	31.087	.000	1.33	353.51	27.96
	N. Words	279.881	33.922	.000	1.38	1856.93	144.94
Fluency	Art. rate	217.650	4.140	.022	0.52	.00	.00
	N. filled pauses	217.374	.529	.234		.02	.00
	M. length pauses	218.788	.018	.446		.02	.00
	N. repetitions	217.135	.056	.407		.00	.00
	N. repairs	216.132	1.372	.122		.00	.00
Syntactic complexity	Global complexity	196.668	2.099	.075		13.64	.30
	Subordination	195.500	.284	.298		.70	.05
	Length	195.005	.082	.388		.78	.07
Lexical complexity	MTLD	196.592	2.163	.072		282.59	2
	Word frequency	201	3.606	.030	0.58	.02	.00
	Word length	201	.100	.377		.09	.00
Accuracy	Error-free clauses	194.296	5.494	.020	0.71	.02	.00
	Lexical errors	194.987	4.797	.015	0.65	.95	.03
	Syntactic errors	195.943	.373	.271		.83	.00
	Morphological errors	194.713	.754	.193		3.31	.05
	Preposition errors	200	.035	.426		.63	.00
Cohesion	Transition markers	194.511	.031	.430		5.10	.20
	Frame markers	194.886	15.517	.000	1.17	.69	.02
	Gloss markers	200	2.709	.051		.52	.00
	Conclusion markers	200	5.344	.011	0.70	.06	.00
	Diversity type	200	20.513	.000	1.38	.38	.00
	Diversity token	200	12.706	.000	1.08	2	.00

Numerator df = 1

The results show that the intervention students made a significant improvement in both quantity measures: quantity of speech ($F(1, 277.886) = 31.1$, $p < .001$) and number of words produced ($F(1, 279.881) = 33.9$, $p < .001$), with large effect sizes.

With respect to fluency measures, the intervention students showed a significant improvement in inverse articulation rate ($F(1, 217.650) = 4.1$, $p = .022$), with a moderate ES (.52). The other fluency measures did not reach the level of significance.

As to syntactic complexity, no significant differences were observed between the two groups in the posttest. However, as to lexical complexity, the intervention group significantly outperformed the control group in terms of word frequency ($F(1, 201) = 3.6$, $p = .030$), with a moderate ES (.58). The other two lexical complexity measures, however, fell short of significance.

With regard to accuracy, the intervention students made fewer errors after the intervention relative to the control students. We found significant differences for error-free clauses ($F(1, 194.296) = 5.5$, $p = .020$), with a moderate to large ES (.71) and for lexical errors ($F(1, 194.987) = 4.8$, $p = .015$), again with moderate to large ES (.65). No significant differences were found for morphological, syntactic, and preposition errors. Note that the distribution of error types did not change greatly over the intervention period.

MLM analyses also revealed that the intervention students showed a significant improvement in many aspects of cohesion compared to the control students. They made significant gains in terms of the use of frame markers ($F(1, 194.886) = 15.5$, $p < .001$), with a large ES (1.17), conclusion markers ($F(1, 200) = 5.3$, $p = .011$), with a moderate to large ES (.70), diversity of type ($F(1, 200) = 20.5$, $p < .001$), with a large ES (1.38), and diversity of token ($F(1, 200) = 12.7$, $p < .001$), again with a large ES (1.08).

6.7 Discussion

The goal of the current research was to provide insights into the effects of debating as an instructional approach on L2 oral proficiency. The main finding of the study was that the debate group scored higher relative to the control group in most variables tested, with many differences reaching significance. These results generally confirm our hypothesis and are in line with previous related studies that suggest that debate-based instruction leads to improved speaking performance (e.g., El Majidi et al., 2018; Zare & Othman, 2015; see Chapter 5). However, these studies were based on debaters' perceptions and instructors' impressions, lacked control groups, and were conducted in higher education. Therefore, the current study provides empirical evidence that substantiates these claims. It is noteworthy that in a recent empirical study (El Majidi et al., 2020; see Chapter 7) investigating the effects of debate pedagogy on L2 writing proficiency, we reported results that, broadly speaking, are comparable to the effects obtained in this study. Though that study assessed effects on writing competence, its comparable findings bolster extra

support for the reliability of the results ensuing from the current study. We now discuss our findings with regard to each measure.

Speech quantity. The intervention students showed substantial growth in quantitative measures of oral performance. The notable increase in overall productivity is an indicator of progress in their language ability, as "students [become] more certain of their skills and [have] more to say" (Lys, 2013). This improvement also attests to the development of the argumentative competence of the debaters. See El Majidi et al. (2021; see Chapter 8) for extensive coverage and discussion of this aspect.

Fluency. The participants in the intervention group showed a significant improvement in the speed measure of fluency. Some of the features of the intervention seem to promote fluency. For example, the "performance aspect of actually doing something in real time" (Schmidt, 1992, p. 359), emphasis on meaning-making (Gatbonton & Segalowitz, 2005), pre-task planning (e.g., Ellis, 2003), task repetition (e.g., Goh, 2017), and close interconnection between the written and oral modalities (Blake, 2009) are all argued to promote fluency. In addition, the intervention enabled sustained practice that has apparently contributed to the learners' development of a degree of automatization in their performance. Improvement in vocabulary also seems to increase the pace of their oral production. When appropriate lexical chunks are readily available, fewer searches are needed, therefore accelerating the formulation process, resulting in greater fluidity in oral production.

Other aspects of fluency—breakdown fluency and repair fluency—did not improve. Earlier research has demonstrated that development in breakdown fluency (e.g., Tavakoli et al., 2016) and repair fluency (e.g., Huensch & Tracy-Ventura, 2017) is less sensitive to interventions and that these aspects of fluency do not necessarily reflect L2 proficiency (Duran-Karaoz & Tavakoli, 2020). Measures of pausing (De Jong, 2018; Duran-Karaoz & Tavakoli, 2020) and repair (Baker-Smemoe et al., 2014; Duran-Karaoz & Tavakoli, 2020; Huensch & Tracy-Ventura, 2017) reflect more personal speaking styles carried over from L1 than L2 proficiency. According to De Jong (2018), "articulation rate (or its inverse average syllable duration) seems to be a measure of fluency that best reflects L2-specific fluency" (p. 250).

Syntactic complexity. The debate intervention did not lead to more syntactic complexity in the produced discourse. However, our recent study (El Majidi et al., 2020; see Chapter 7) demonstrated that debate pedagogy exerted some impact on the syntactic complexity of written production. Hwang et al. (2020) suggested that L2 learners "experience greater processing burdens in spoken than in written production" (p. 272), and this hinders them from producing syntactically complex structures in their spoken production. The lack of improvement in

syntactic complexity could also be due to a trade-off effect, as the debaters could have focused in their oral production on fluency, accuracy, and lexical complexity at the expense of syntactic complexity (Skehan, 1998). The lack of improvement in syntactic complexity because of a possible trade-off effect is consistent with the findings of previous studies (e.g., Hsu, 2019).

Lexical complexity. The intervention group outperformed the control group in two measures of lexical complexity: measure of textual lexical diversity (MTLD) and word frequency, with the latter reaching significance. MTLD and word frequency have been regarded as important indicators of lexical proficiency, as learners who produce less frequent (Crossley, Salsbury, et al., 2015) and diverse vocabulary (Crossley et al., 2009) are judged to be more lexically proficient. Word frequency measure is also associated with breadth of lexical knowledge, as learners who produce infrequent words are expected to have knowledge of a greater number of words.

The positive effect of in-class debates on lexical complexity might be attributable to the lexically rich environment of debates. The students read articles prior to each debate. By means of reading, the participants seemingly gained access to words and structures of which they were unaware. Importantly, case writing created an opportunity to employ the newly learned words and hence consolidated their retention. During actual debates, the debaters also used some of these words, a fact that further reinforced their grip on these lexical gains. This cyclic lexical process, which recurred during each debate, enabled the debaters to incrementally build an extensive, diverse, and sophisticated vocabulary.

Accuracy. The debate intervention seems to yield some benefits with respect to accuracy. Although the learners in both groups showed some improvement, the intervention group made more significant improvements in two of the accuracy indices. The greater improvement made by the intervention group may have resulted from their exposure to feedback in the context of the debate-based lessons. The intervention group received feedback on their written cases, and during debates they were instructed to note down the mistakes their classmates made and improve them. In addition, the instructors occasionally discussed some of the mistakes commonly made during debates. It seems that the debaters benefited from the recursive cyclic processing of feedback. This cyclic processing of feedback, which recurred in each debate, enabled the debaters to carry over the accuracy gains from one debate to the next. In this way, gains piled up and led to better monitoring (processes) of online speech production. Furthermore, the combination of the written and oral production modalities could have heightened the debaters' attention to language forms (see Niu, 2009).

The debate environment also seems to facilitate a meaningful processing of feedback, as accurate language is essential for the persuasiveness of the adduced arguments. In other words,

accurate form in this environment is functional. Furthermore, we believe that the presence of an audience (proponents and critical opponents) prompted the students to pay extra attention to their language production to avoid making embarrassing grammatical mistakes. Research has revealed that the sense of audience awareness stimulates learners to attend to the quality of their language output (Chen, 2019).

Cohesion. The current study has also tracked the effects of the debate intervention on a set of organizational markers, which explicitly signal structural organization within an oral discourse. These organizational markers contribute to crafting a coherent and interrelated discourse by connecting speakers' arguments to each other.

In the posttest, the debaters produced speech samples that demonstrated significant growth in terms of different cohesion areas. It is important to mention that the students in the intervention group were asked to use connectives in their written cases during the intervention (see Chapter 3.4.2). Apparently, via this channel organizational markers found their way to the oral discourse. Importantly, the debate discourse prompts debaters to explicitly organize their discourse and highlight the relationship between the arguments presented. For example, adducing different arguments entails the use of transition and frame markers to signal progression and a shift from one argument to another. The debate environment presumably infused into the debaters the conviction that making explicit relations between arguments is essential for their discourse to be persuasive. What is more, the engaging and challenging nature of the debate environment may also have contributed to the enhancement of this performance area (Cinganotto, 2019; Crosson & Lesaux, 2013). From this, we can conclude that the debate context seems to afford fertile ground for the development of organizational markers.

6.7.1 Factors reinvigorating debate pedagogy

In this section, we discuss the main factors that seemingly render debate pedagogy effective for L2 speaking development. It could be argued that the intervention students in the present study have developed their speaking performance more simply due to the extra amount of practice time they received. Though the control group also practiced speaking, the intervention group, quantitatively speaking, practiced more, and this may have honed their oral performance. However, although we acknowledge that there is a possible 'practice' effect, we argue that this effect has not solely led to the emerging findings (see Lys, 2013). There are many relevant intervention-related factors that have arguably given rise to the findings of the current study.

The debate task design and environment possess features and involve procedures that theoretically and empirically were proved to affect the processes of complex, accurate, coherent, and fluent production. The debate task design involves, for example, task repetition. Task

repetition "involves the repetition of the exact same task but may also be a repetition of the same task procedure with different contents" (Goh, 2017, p. 249). In our intervention, students (at least many of them) rehearsed their prepared cases at home prior to their delivery in class, and the debate task procedures were repeated with "different content" (different debate topics) throughout the whole intervention. When students repeat a task (procedure), they free up attentional resources and hence boost their performance by facilitating a speedy retrieval of words and integrating knowledge and skills gained from the first performance. As a result, partial automatization of some speech processes takes place, thereby directing attention to other aspects of speech (Bygate, 1996; Goh, 2017).

Another feature of our debate task design, which has possibly led to the obtained results, is the integration and synergy of the four language skills that have presumably scaffolded oral production in many ways. Prior to each debate contest, the students read preparatory articles on which they drew to write cases and which, in turn, fed oral production during debates. The integration of these skills—or rather reading-to-write-to-speak pedagogy (where reading is used as input for writing, and writing is used as input for speaking)—substantially boosts speaking. The combination of reading, writing, and speaking is assumed to engineer a powerful pedagogical configuration that draws more attention to language forms than oral production by itself, and it offers great language learning opportunities.

We are not aware of an earlier study examining the potential of teaching tools that combine these three skills. However, evidence that points to this pedagogical formula leading to positive learning outcomes can be adduced from reading-to-write and writing-to-speak pedagogies. Hirvela (2016) has pointed out that "reading… provides writers with essential material to write with and about" (p. 49). It seems that the learners profited from the recurrent process of reading and writing, which allowed them to employ the newly learned words in the subsequent cases. Earlier research has demonstrated the effectiveness of L2 reading-to-write pedagogy, as extensive reading helps students to improve their vocabulary (lexical complexity) and to write better (e.g., Hirvela, 2016).

As to writing-to-speak pedagogy, Belcher and Hirvela (2008) stated that writing involves a slow cognitive processing of language that promotes explicit noticing, and through repeated use of explicit knowledge, implicit knowledge incrementally increases; this results in an increased automatic oral fluency and accuracy (Adams & Ross-Feldman, 2008). In addition, "writing activities before speaking activities … allow learners to apply features of their written production to their spoken production [and help] learners efficiently retrieve more varied and complex syntactic structures from their writing experience" (Hwang et al., 2020, p. 279). Also, because writing allows offline planning, it stimulates learners to try out new forms (Williams, 2012) and facilitates increased monitoring of output (Ellis & Yuan, 2004). As such, writing

allows working out and reflecting upon the intended message before delivering it in spoken interaction. Having used a particular form in planned writing enhances its retention, and thereby increases the likelihood of its easy retrieval and use in unplanned spontaneous oral production (Williams, 2012). In this way, gains in writing transfer to speaking skills. The importance of writing for fluidity of spoken language has been documented in a number of studies (Freed et al., 2004; Rubin & Kang, 2008).

In addition, the interactions promoted by the debate environment harness listening to serve speaking. Through critically listening to the output of their peers, debaters notice gaps in their output. This awareness can induce learners to scan future input to find remedies for these language gaps (Williams, 2012).

6.8 Limitations and future directions

Our study has some limitations that are worth noting. The findings in this study were based on a single task type (i.e., opinion tasks). Since performance tends to vary across task types and genres, future studies should consider employing a wider variety of different tasks to gain deeper insights into the observed effects and their transferability. In addition, this study elicited data from two time points (pretest and posttest). A delayed posttest (i.e., a third time point) would provide insights into the durability of the performance patterns observed.

The present research has not examined the impact of debating on pronunciation, which is an important dimension of speaking performance. Its clarity, for example, is deemed to be an important feature of comprehensibility (Suzuki & Kormos, 2020). Some debaters in a previous study (El Majidi et al., 2018; see Chapter 5) reported that they felt that their pronunciation had improved after participating in debates. Future research may study the impact of debating on this construct to provide a more complete picture of the impact of debate pedagogy on speaking proficiency. Last but not least, our sample is not representative of all secondary school students, as it mainly includes relatively higher-achieving students. More research with a wider range of language proficiency groups is warranted to further broaden our understanding of the potential of debate-based instruction for L2 speaking development.

6.9 Conclusion

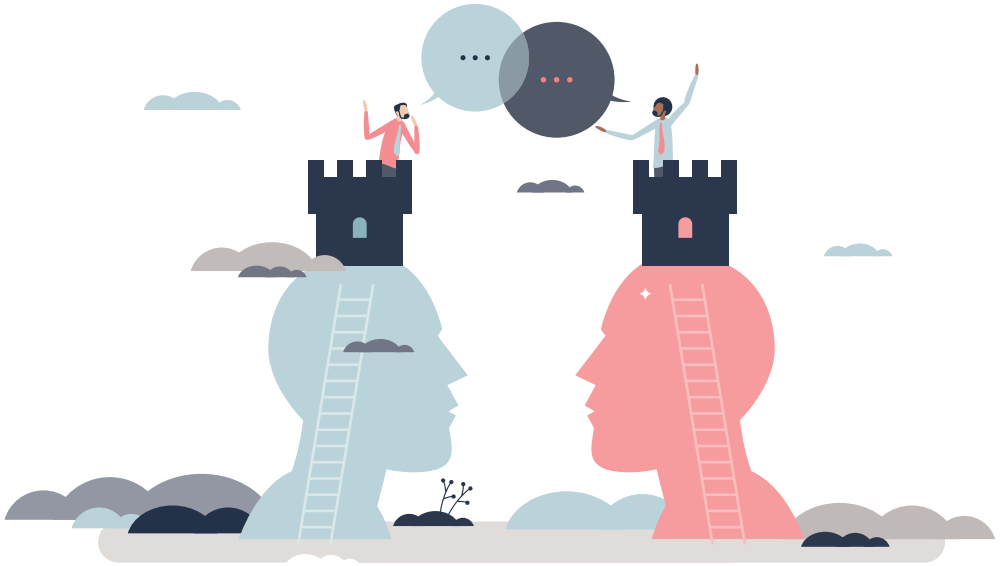
The present study has offered empirical evidence that debate can play a facilitative role in enhancing L2 speaking proficiency. After the intervention, the debating students produced significantly more words delivered at a faster pace and produced a speech that is more accurate, lexically sophisticated, and coherent than the students in the control group.

These gains could be attributed to the iterative process of successive waves of input and output the debaters went through as well as to the deeper language processing in which the debate context engaged the debaters. Debate pedagogy seems to trigger deeper and more elaborate processing of content and form, which helped the students to achieve greater proficiency and automaticity over time. This depth of processing is particularly promoted by the reading-to-write-to-speak pedagogy facilitated by the debate task design. This pedagogy enables a synergetic and harmonious collaboration of the language skills, which lead to exposing learners to rich and authentic L2 input, and provides socially oriented opportunities for meaningful, authentic, purposeful, and goal-oriented interaction.

We hope that the potential of debate for enhancing speaking proficiency in instructed L2 acquisition will be recognized and will hence be included in L2 curricula. We also hope that we have made some modest contribution to the current discussion about teaching speaking through providing insights into some processes and task features that stimulate and scaffold L2 oral proficiency.

Chapter 7

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



This chapter is based upon: El Majidi, A., de Graaff, R., & Janssen, D. (2020). Debate as L2 pedagogy: The effects of debating on writing development in secondary education. *The Modern Language Journal, 104*(4), 804–821. https://doi.org/10.1111/modl.12673

Abstract

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 effects 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 eight classes at three secondary schools in the Netherlands ($N = 146$). To measure the effects of the intervention, we analyzed two 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

7.1 Introduction

Writing effectively is regarded as a vital twenty-first 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 et al., 2010), and 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). Given 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 et al., 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 effects 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.

7.2 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

[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.

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) contended 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 Section 7.5.3), 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 facilitate 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 et al., 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).

7.3 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 et al. (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 instructors 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 modality 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 modalities 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 for the produced content—namely, argumentative skills (e.g., Chen, Hand, et al., 2016; Chen, Park, et al., 2016).

In the debate context, composition is viewed as a socially oriented activity: Participants produce cases for a particular purpose and are 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 towards debating as a pedagogical tool (e.g., El Majidi et al., 2015b, 2018; Lustigova, 2011; see Chapters 4 and 5). 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).

7.4 The present study

Very few studies in either the first language or the L2 context have examined the effects of debate on language performance in general and on writing performance in particular. The studies that have explored these effects in L2 or FL so far were based on observations (Lustigova, 2011), interviews (Aclan & Aziz, 2015a), or questionnaires (mainly focused on speaking; Zare & Othman, 2015); and 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; see Chapter 5). The participants in this study perceptively correlated debate participation with an improvement in writing with a mean of 3.94 on a five-point Likert scale.

To the best of our knowledge, no empirical studies have been conducted in L2 or FL research to demonstrate the extent to which in-class debates improve students' writing proficiency across different areas of performance. Polio and Park (2016) called for charting the

effects of interventions on different linguistic aspects of writing, 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 addressed the following research question:

RQ.1 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 of Dutch secondary school students?

H.1 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 (posttest).

7.5 Method

7.5.1 Participants

The study sample consisted of eight intact classes at three secondary schools in the Netherlands ($N = 146$) where English is taught as an L2. These schools are located in three urban areas (Rotterdam, Leiden, and Alkmaar) with comparable graduation rates. Five classes were in the fifth year of higher general secondary education track (*havo 5*)[2] ($n = 88$), and three classes were in the fourth year of the preuniversity secondary 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

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

(the fourth level of English in CEFR; this level is comparable to the upper intermediate level) levels 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 purpose 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 course books, which mostly require students to produce short texts.

7.5.2 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.

7.5.3 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 one 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 Chapter 3.4.2).

Each debate consisted of three stages: pre-debate, debate, and post-debate (see Appendix B). Table 7.1 presents the writing activities performed in each debate session, in addition to the writing tasks performed by the control group during the intervention.

Table 7.1 Relevant activities performed during the intervention

Intervention group	Control group
Pre-debate stage - Reading two articles and summarizing them. The instructor provided one article and the students had to find another one - Writing a case **During-debate stage** - 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 **Post-debate stage** - Processing the feedback provided by the instructor on written cases	- 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 - Writing persuasive essays on controversial subjects (e.g., school uniforms should be introduced in schools) - Writing letters, especially complaint letters in which students had to express dissatisfaction with a particular service and accordingly convince the addressed company to provide a refund

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.

7.5.4 Procedure

To measure the effects of our debate intervention on writing proficiency, we compared two free opinion tasks (as pretest and posttest). 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 pretest–posttest (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 the control group). The conditions around pre- and post-assessments were the same for both the intervention and control groups. In both groups, students did not receive a grade for these tasks.

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 learners' 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.

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.

7.5.5 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 et al., 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 et al., 2016), and it 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:

- 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 analyzer (Lu, 2010), which was specifically developed to parse L2 written data.

[3] Dobbs (2014) used persuasive tasks to refer to opinion tasks.

Accuracy. To measure accuracy, we first segmented the written texts (292 in total) into clauses. Following Miller's (2008) guidelines, 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:

- Error-free clauses (EFCs)
- Lexical errors per 100 words
- Syntactic errors per 100 words
- Morphological errors per 100 words
- Prepositional errors per 100 words

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 and punctuation errors, unless a misspelled word resulted in an actual English word (Ferris & Roberts, 2001).

Lexical complexity. To track the effects 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 et al., 2010):

- Average word length (WL)
- Word frequency (WF)
- 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 et al., 1995). A lower WF thus indicates higher sophistication. We used MLTD to measure lexical diversity. In addition to the fact that MTLD

[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.

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, p. 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 to assess cohesion. This framework assesses interactive metadiscourse markers, which are words and phrases used to explicitly signal the coherent organization of ideas and arguments in a text to guide readers. Coh-Metrix also provides 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 markers, in addition to their diversity of type and token (see, for example, Dobbs, 2014; Qin & Uccelli, 2016; Uccelli et al., 2013):

- Frame markers: markers that mark the sequence of arguments or counterarguments (e.g., firstly, secondly)
- Code glosses markers: markers that introduce an example or paraphrase (e.g., for instance, in other words)
- 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., in conclusion, all in all)

- Markers diversity token: diversity of markers in terms of token
- Markers diversity type: diversity of markers in terms of type

7.5.6 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.

7.5.7 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 model analyses (MLM). 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 modeled the independent variables (time and condition) as fixed effects, and random variations across students and classes as random effects.

To establish the effectiveness of the debate intervention, we need to take into account the combined effect of both main factors. In other words, we need to focus on the interaction of time (pretest 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 through a visual analysis of the histograms of each residual, which is the standard procedure in multilevel modeling. No notable deviations were visible.

7.6 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 E for an example of the analysis of a text produced by a participant). Table 7.2 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 in the posttest.

Table 7.2 Means and standard errors across time and condition

Measures	Index	Intervention (n = 95)				Control (n = 51)			
		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	Error-free clauses	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.

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 MLM results shown in Table 7.3 are presented in relation to the measures discussed in the previous section.

Table 7.3 Multilevel analysis results

Measures	Index	Fixed effects				Random effects	
		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.

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 than in the pretest, while the control participants produced six words more in the 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.

With regard to 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 posttest (see Table 7.2); however, this improvement fell short of significance.

With respect to accuracy, we found that the debate intervention affected various aspects of accuracy to different degrees. Although learners in both groups displayed an improvement across almost all accuracy measures, MLM 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 a 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, MLM 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 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 large effect size, $d = 1.27$; and diversity of token, $F(1,284.018) = 37.5$, $p < .001$, again with a 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 in the posttest as Table 7.4 displays.

Table 7.4 Frequency of the use of sophisticated connectives

Connectives	Control ($n = 51$)		Intervention ($n = 95$)	
	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

7.7 Discussion

The main goal of the present study was to assess the impact of an in-class debate 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 the 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 that 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 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 the findings in more detail.

Fluency. The participants in the intervention group showed a 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, 2007a).

Note also that the intervention students were instructed to read at least two argumentative articles prior to each debate. It seems that these students 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, showing 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 7.2). 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 that captures syntactic complexity development (e.g., Lu, 2010; Norris & Ortega, 2009). MLT and CT measures did not reach significance. The mean of MLT produced in the pretest and posttest for both groups did not change greatly. This finding is in line with the findings of previous studies (e.g., Knoch et al., 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 in the 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, et al., 2015) 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; see Chapter 5) in which the debaters reported improvement to their vocabulary after participating in L2 debates.

Cohesion. Beyond the well-known areas of fluency, lexical and syntactic complexity, and accuracy, 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 an 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] (see Chapter 3.4.2). More importantly, the debate context seems to provide a natural and fertile environment 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).

7.8 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 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.

[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 the pretest and posttest.

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 facilitates 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 towards 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 three 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

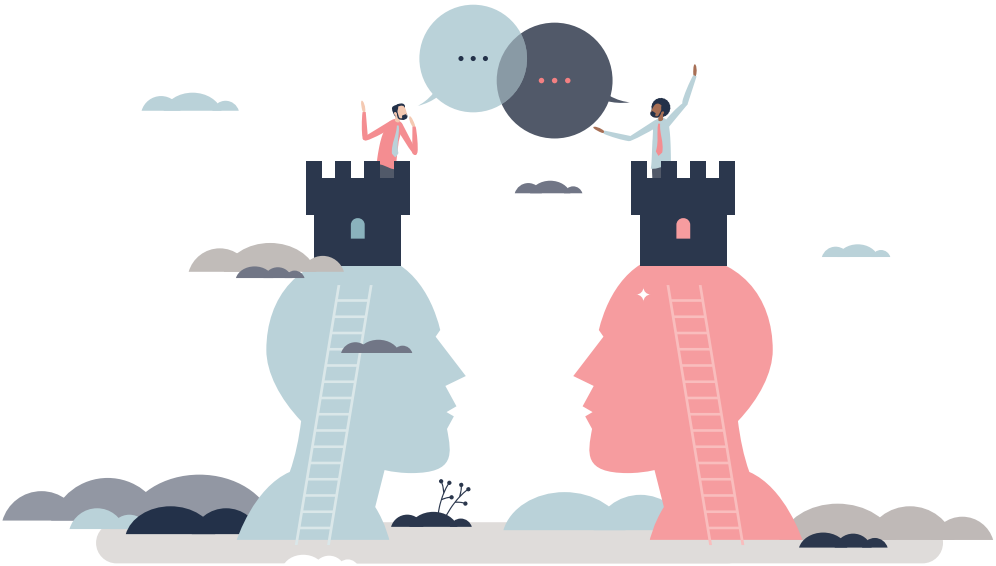
measure the effects of debate-based instruction on writing proficiency and confirm these findings.

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 the extent to which 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 (pretest and posttest). Future studies would also benefit from a delayed posttest time point. This third time point would yield insights into the durability of the potential effects and hence provide additional evidence for the performance patterns observed.

Chapter 8

The Effects of In-Class Debates on Argumentation Skills in Second Language Education



This chapter is based upon: El Majidi, A., Janssen, D., & de Graaff, R. (2021). The effects of in-class debates on argumentation skills in second language education. *System*, 101.https://doi.org/10.1016/j.system.2021.102576

Abstract

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

Keywords: in-class debates, second language, argumentation skills, secondary education, written and oral argumentation, argumentation pedagogy

8.1 Introduction

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

Many studies have voiced concerns about students' argumentative abilities in the first language (L1) (e.g., Chen, Hand, et al., 2016; Crowell & Kuhn, 2014) as well as in the second language (L2) (e.g., Huh & Lee, 2014; Qin & Karabacak, 2010). Empirical evidence shows that students struggle with providing adequate justifications for their claims (Hsu et al., 2015) and generating counterarguments to rebut the opposing side (Liu & Stapleton, 2020; Sadler, 2004; Stapleton & Wu, 2015). Weaknesses in argumentative reasoning can have serious implications for success in academic and professional careers and therefore this warrants nurturing (Crowell & Kuhn, 2014; Kuhn, 1991; Kuhn & Udell, 2003; Yeh, 1998). Or, in the words of Yeh (1998), the capacity to produce "effective arguments influences grades, academic success, and preparation for college and employment" (p. 49).

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

There are two notable variables that can interfere with argumentation performance in L2: culture and language proficiency. Culture is a variable that may play a role in the development of argumentation skills. In some non-Western societies like countries dominated by Confucianism culture, teachers are viewed as authority figures, who cannot be challenged or criticized by

students (Heimgärtner, 2013). However, in most Western countries (the context of this study), which are characterized by small power distance societies, as teachers and students relatively share more equal power distribution, students are stimulated to speak up and can even challenge teachers (Hsu et al., 2017). In such a context, the culture variable is less likely to interfere with argumentation expression and development.

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

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

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

Though there is wide recognition that debate may promote argumentation skills, there is no empirical evidence available yet to support this assumption in the L2 context. To fill this gap,

we conducted an intervention study in which we investigated the effects of in-class debates on the development of argumentation skills of Dutch secondary school students in English classes by identifying and comparing developmental features (structural and quality aspects) of argumentative skills in samples of their written and oral production.

8.2 Theoretical grounds

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

Another theoretical perspective has to do with the conceptualization of argumentation: *learning to argue* and *arguing to learn* (Hirvela, 2017). In the *learning to argue* conceptualization, students are taught how to argue through teaching them the components of argumentation; in other words, arguing is seen as an end in itself. On the other hand, the *arguing to learn* perspective conceptualizes argumentation as a "means by which we rationally resolve questions, issues, and disputes and solve problems" (Jonassen & Kim, 2010, p. 439). According to Hirvela (2017), the *arguing to learn* orientation empowers learners to think beyond the argument structure and fosters their analytical and critical thinking skills. Debating may subscribe to *the arguing to learn* orientation. In debates, argument is perceived as a medium through which a

functional goal can be attained, and that is, defending one's position and weakening that of their opponents. This orientation stimulates students to see argument as a "process, not product" (Hirvela, 2017, p. 72).

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

The audience-centered approach to argumentation that characterizes debate presents another relevant perspective that supports our hypothesis. Many studies showed that audience awareness pushes debaters to hone their reasoning skills. For example, Chen, Hand, et al. (2016) pointed out that when "students [are] expected to be challenged and critiqued by their peers [they push] themselves to refine and reshape their argument to be convincing" (p. 130). In the same vein, Midgette et al.'s (2008) study revealed that students with audience-oriented goals were more likely to consider opposing positions and rebut them. Even during preparation (e.g., case[1] writing), audience awareness can affect the construction of arguments, in that arguers consciously think about the audience's objection and response and accordingly reflect upon and refine their line of reasoning (Johnson, 2013). The last theoretical perspective has to do with students' attitude towards debate. Students' favorable attitude towards debating is likely to lead to high cognitive engagement in the learning process, thereby leveraging the learning potential offered by the debate experience (El Majidi et al., 2015b, 2018; see Chapters 4 and 5).

[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).

8.3 Debate-argumentation research

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

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

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

Hirvela (2017) regarded neglecting argument in the L2 research agenda (especially in L2 writing) as odd, unjustifiable, and unacceptable, given its manifest importance. For example, L2 argumentative competence is viewed as an important indicator of L2 writing ability "as argumentation is at the heart of SL [second language] writing assessment" (Hirvela, 2017, p. 69). In addition, L2 argumentation may also boost L2 language development. Chapple and Curtis (2000) stated that improvement in L2 analytic and critical thinking may lead to improvement in L2 language skills. Similarly, Pally (1997) contended that "complex, synthetic

reasoning needs to be practiced in the L2 in order for students to master—and challenge—L2 language and argument" (p. 299).

8.4 This study

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

RQ.1 What are the effects of in-class debates on the structural components of the written and oral arguments produced by Dutch secondary school students?

RQ.2 What are the effects of in-class debates on the quality of the written and oral arguments produced by Dutch secondary school students?

The theoretical considerations (discussed above) and previous research (e.g., Oros, 2007; Zorwick & Wade, 2016) led us to anticipate:

H.1 that the students engaged in the debate intervention would produce more written and oral arguments and that these arguments would include more sophisticated structural patterns.

H.2 that the students engaged in the debate intervention would produce more written and oral arguments that exhibit better reasoning quality.

8.5 Method

8.5.1 Design

To answer these research questions, an intervention with a pretest–posttest control group quasi-experimental design was conducted. The intervention group participated in 10 debates (one debate per week), which were part of the class curriculum, with each lasting approximately 50 minutes. It is noteworthy that the intervention group did not receive instruction about the structural and quality aspects of argumentation. To enhance the external validity of the findings, we gathered data from three different secondary schools. The participating teachers were instructed and familiarized with the content of the intervention by the first author. The data for this study consisted of written and oral opinion tasks—in which the intervention and control students argued for/against a side of a controversial topic; see Section 8.5.4 for more elaboration—that were elicited at the beginning (pretest) and towards the end (posttest) of the intervention.

We believe that a free writing opinion task, which is not governed by any particular writing conventions, would allow students to focus more on arguments and their articulation[2]. Furthermore, we believe that free opinion tasks are likely to provide a more fine-grained picture of students' use of counterarguments and rebuttals than essays since our students were already instructed (before the intervention) to compose two-sided essays.

8.5.2 Participants

The study's sample consisted of eight intact classes at three secondary schools in the Netherlands ($N = 147$). Five classes were in the fifth year of higher general secondary education ($n = 89$), and three classes were in the fourth year of preuniversity secondary education[3] ($n = 58$). Five classes served as the intervention group ($n = 96$) and three as the control group ($n = 51$). The participants included 88 females and 59 males, ranging in age from 15 to 18. The English proficiency level of all classes spanned mostly the B1[4] and B2[5] levels, 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 classes received three sessions of 50 minutes. Both groups

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

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

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

[5] Upper-intermediate level according to the CEFR.

received regular instruction consisting of activities dealing with the four language skills. For the purpose of this study, while the intervention students were involved in the debate intervention (once a week), the control students received extra regular instruction (in which the four skills were further practiced).

8.5.3 The intervention

The debate task was developed and validated in a previous study following the principles and guidelines of educational design research (see, for example, McKenney & Reeves, 2012). The intervention students, who participated in 10 debates, were informed at least one week in advance of the debate topics. They were allowed to choose the topics that interested them and the side they wanted to defend (i.e., an affirmative or a negative side). We employed two debate formats that were extensively tested before the intervention took place: debating in a group of four debaters (two students in favor and two against) and a one-to-one debating format. All debates had three phases: constructive speech, rebuttal, and clash (see Chapter 3).

Table 8.1 presents the activities performed in each debate session, in addition to the tasks involving argumentation performed by the control students during the intervention.

Table 8.1 Main activities conducted during the intervention.

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

Since students in the intervention group (in different schools) were allowed to choose topics, their readings were not exactly the same. This also applied to the students in the control group, who were allowed to choose their own articles (which, as indicated in Table 8.1, included argumentative articles). It follows, then, that specific content is not likely to interfere with our findings. It is worth mentioning that the intervention students also wrote persuasive essays and letters. However, in order not to disadvantage the control students in terms of the amount of practice with the argumentative discourse, we allowed them to write more essays and letters. More precisely, during the intervention the control group wrote on average two more essays

and one more letter than the intervention group. Additionally, the control groups read and summarized three more articles, some of which were argumentative.

8.5.4 Procedure

To examine whether the debate intervention had an impact on written and oral arguments, we analyzed and compared two written and two oral texts (argumentative opinion tasks) produced by the intervention and control groups during the pretest and posttest (see Table 8.2). As to the writing task, we selected two controversial topics: (1) capital punishment should be legalized; and (2) abortion should be banned. Topics like capital punishment and abortion compared to many other social issues are accessible topics with clear sides (with each side having abundant arguments). Having clear sides was likely to enable our participants to generate counter-arguments and rebuttals. These topics were previously piloted and proved to fit our context.

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

Table 8.2 Data collection procedures

Pretest	
- Writing a 15-minute opinion task after 25 minutes' preparation time - Producing an oral opinion (with no time limits) after seven minutes of preparation time	
Intervention (intervention group)	No intervention (control group)
See Table 8.1 for the activities conducted during the intervention	See Table 8.1 for the activities conducted during the intervention
Posttest	
- Writing a 15-minute opinion task after 25 minutes' preparation time - Producing an oral opinion (with no time limits) after seven minutes of preparation time	

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

15-minute timed text defending their side. The topics of the writing pretest and posttest (capital punishment and abortion) were counterbalanced to avoid any potential topic effect.

8.5.5 Data analysis

Argument structure analysis

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

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

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

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

In the current study (see Appendix F), we coded Toulmin's primary elements (i.e., claim, data, and warrant), and following Crammond (1998) and Cheng and Chen (2009), we have added four structural elaborations and modifications to Toulmin's model. Firstly, the backing element in the present study includes, in addition to backing for warrants, backing for data, qualifiers, alternative solutions, counterarguments, and rebuttals. Secondly, qualifiers are modified to include the conditions that limit the applicability or validity of the claim (reservation),

or the conditions under which the claim would apply (constraints). Thirdly, unlike in the Toulmin model, rebuttals in the present study refer to the debater's response to opposing views that challenge the debater's claim. Fourthly, following Crammond, and Cheng and Chen, we also coded alternative solutions that refer to possible solutions or answers to the issue under consideration. Lastly, after we had performed some preliminary coding, we realized that it was necessary to include an additional element also employed by Karabacak and Qin (2016), namely background information, which consists of general information about the issue, and to introduce a novel element into our analysis, i.e., subordinate arguments, which are subarguments that back up a main argument.

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

To perform structural analysis, all texts received an overall score for argumentation complexity and a frequency score for each dependent variable (e.g., data and subarguments). For example, the subargument score reflects the number of subarguments used in each text.

Argument quality analysis

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

To assess the quality of each production, we developed a five-point rating scale comprising a set of items that tap into different aspects of the reasoning quality, including organization (e.g., the extent to which the arguments are well organized and flow well), sufficiency, clarity, elaboration, relevance, persuasiveness, and addressing the opposing view, in addition to holistically assessing the overall quality of each production (see Appendix H for more elaboration). Following this model, arguments increase in quality as they become increasingly organized, sufficient, clear, elaborated, relevant, persuasive, and address the opposing view. In developing this quality-oriented evaluation framework, we drew heavily on coding schemes developed and

validated by Qin and Karabacak (2010) and Ferretti et al. (2000). In Appendix I, we have provided an example of the scoring procedures for two written texts.

8.5.6 Interrater reliability

To perform a structural analysis of the written and oral arguments produced, we identified elements of arguments and calculated their frequencies. Efforts to achieve acceptable interrater reliability were made as follows: The first two authors first went through practice sessions, scoring some texts, and then compared scores and discussed the discrepancies to clarify and standardize/calibrate the interpretation of coding. After the practice sessions, the first author coded the entire data set. To calculate reliability scores, a random sample of 25% was checked by the second author, who was not involved in carrying out the intervention and data collection and was blind to condition. Reliability scores were calculated using Cohen's kappa and were as follows: background information (1), claim (1), data (.83), subarguments (.81), warrants (.82), counterarguments (.92), rebuttals (.93), qualifiers (.93), alternative solutions (.97), and backings (.92).

The first author and a research assistant coded the quality aspects of the data. After the practice sessions, they independently scored a randomly selected sample of 25% of the data (of both written and oral arguments). Reliability scores were calculated using Cronbach's alpha and were: overall assessment (.87), organization of arguments (.82), sufficiency of arguments (.89), clarity of arguments (.81), elaboration of arguments (.86), relevance of arguments (.77), persuasiveness of arguments (.79), and addressing opposing views (.61). After resolving the disagreements, the research assistant who was unaware of the data sources scored the remaining data.

8.5.7 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. In other words, we need to focus on the interaction of time (pretest vs. posttest) × group (intervention vs. control group). For these reasons, we limit ourselves to reporting interactions.

8.6 Results

8.6.1 Structural analysis

Table 8.3 presents the descriptive statistics (estimated means and standard errors) for the structural analysis of written data and Table 8.4 for oral data.

Table 8.3 Means and standard errors across time and condition of written data

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

Table 8.4 Means and standard errors across time and condition of oral data

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

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

To answer the first research question about the impact of the debate intervention on the structural elements of the participants' argumentation, we conducted MLM analyses. To estimate the magnitude of the difference between the intervention and control groups, we used Cohen's *d* effect size (ES) when significant differences were observed. The MLM results with regard to structural analysis are presented in Table 8.5 for written data and in Table 8.6 for oral data.

Table 8.5 Multilevel analysis results of the structural analysis of the written data

Measures	Fixed effects				Random effects	
	Denominator df	*F*	*P* (one-tailed)	*d*	Variance within class	Variance between class
Overall argument complexity	284.690	5.610	.010	0.59	1.20	.03
Background information	284.017	.000	1.000		.91	.00
Claim	286.532	1.104	.147		.00	4.31
Data	284.261	4.680	.016	0.51	2.02	.21
Subarguments	284.642	9.105	.002	0.72	2.57	.16
Warrants	284.216	1.606	.103		.24	.00
Counterarguments	281.167	2.893	.045	0.42	.32	.00
Rebuttals	292	3.085	.040	0.43	.37	.00
Qualifiers	292	3.359	.034	0.45	.14	.00
Alternative solutions	292	.360	.549		.19	.00
Backings	284.521	2.993	.043	0.41	.34	.02

Numerator df = 1

Table 8.5 shows that the debate intervention appears to have an impact on a number of structural elements of the participants' written argument. MLM analyses revealed that the intervention group significantly improved in terms of overall argumentation complexity ($F(1, 284.690) = 5.6$, $p = .010$), with a moderate to large ES (.59); the use of data ($F(1, 284.261) = 4.7$, $p = .016$), with a moderate ES (.51); subarguments ($F(1, 284.642) = 9.1$, $p = .002$), with a moderate to large ES (.72); counterarguments ($F(1, 281.167) = 2.9$, $p = .045$), with a moderate ES (.42); and rebuttals ($F(1, 292) = 3.1$, $p = .040$), again with a moderate ES (.43). The students in the intervention group also showed a significant improvement in terms of their use of qualifiers ($F(1, 292) = 3.4$, $p = .034$) and backings ($F(1, 284.521) = 3.0$, $p = .043$), with moderate effects on the two indices: .45 for qualifiers and .41 for backings.

Table 8.6 Multilevel analysis results of the structural analysis of the oral data

Measures	Fixed effects				Random effects	
	Denominator df	*F*	*P* (one-tailed)	*d*	Variance within class	Variance between class
Overall argument complexity	286	9.838	.001	0.78	.83	.00
Background information	291	.219	.320		.01	.00
Claim	291	.537	.232		.00	.00
Data	278.327	22.492	.000	1.14	.90	.04
Subarguments	278.912	12.973	.000	0.89	1.17	.03
Warrants	279.519	6.515	.006	0.65	.15	.00
Counterarguments	286	4.364	.019	0.52	.14	.00
Rebuttals	286	7.369	.004	0.69	.15	.00
Qualifiers	279.997	.007	.466		.01	.00
Alternative solutions	286	.102	.375		.02	.00
Backings	279.421	1.856	.087		.07	.00

Numerator df = 1

As to oral arguments, the students in the intervention group also showed a significant improvement in terms of overall argument complexity ($F(1, 286) = 9.8$, $p = .001$), with a moderate to large ES (.78); the number of data ($F(1, 278.327) = 22.5$, $p < .001$), with a large ES (1.14); subarguments ($F(1, 278.912) = 13$, $p < .001$), again with a large ES (.89); warrants ($F(1, 279.519) = 6.5$, $p = .006$), with a moderate to large ES (.65); counterarguments ($F(1, 286) = 4.4$, $p = .019$), with a moderate ES (.52); and rebuttals ($F(1, 286) = 7.4$, $p = .004$), with a moderate to large ES (.69).

8.6.2 Argument quality analysis

Tables 8.7 and 8.8 display the descriptive statistics of the argument quality analysis for both written and oral data, respectively.

Table 8.7 Means and standard errors across time and condition of written data

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

Table 8.8 Means and standard errors across time and condition of oral data

Measures	Intervention group (n = 96)		Control group (n = 51)	
	Pretest	Posttest	Pretest	Posttest
Overall assessment	2.33 (.06)	2.86 (.06)	2.28 (.08)	2.31 (.08)
Organization of arguments	1.92 (.07)	2.58 (.07)	1.90 (.09)	2.00 (.09)
Sufficiency of arguments	1.89 (.09)	2.60 (.09)	1.98 (.11)	2.02 (.12)
Clarity of arguments	1.98 (.08)	2.64 (.08)	2.18 (.11)	2.08 (.11)
Elaboration of arguments	1.91 (.06)	2.55 (.06)	1.93 (.08)	2.07 (.08)
Relevance of arguments	2.21 (.08)	2.80 (.08)	2.49 (.10)	2.42 (.10)
Persuasiveness of arguments	2.19 (.08)	2.79 (.08)	2.33 (.11)	2.28 (.11)
Addressing the opposing view	1.00 (.03)	1.09 (.03)	1.06 (.04)	1.07 (.04)

To answer the second research question, MLM was performed to establish whether the intervention impacted the quality of the arguments produced by the intervention participants in comparison to their control peers. The results are displayed in Table 8.9 for written data and in Table 8.10 for oral data.

Table 8.9 Multilevel analysis results of the quality analysis of the written data

Measures	Fixed effects				Random effects	
	Denominator df	F	P (one-tailed)	d	Variance within class	Variance between class
Overall assessment	283.955	.672	.207		.38	.06
Organization of arguments	283.887	11.109	.001	0.71	.37	.13
Sufficiency of arguments	284.026	1.173	.140		.52	.17
Clarity of arguments	283.928	4.870	.014	0.49	.50	.11
Elaboration of arguments	284.041	9.155	.002	0.65	.50	.16
Relevance of arguments	283.953	3.937	.024	0.43	.42	.12
Persuasiveness of arguments	284.012	3.289	.036	0.40	.48	.10
Addressing the opposing view	284.925	.905	.171		.35	.01

Numerator df = 1

Tables 8.9 and 8.10 reveal that the debate intervention seems to have improved the majority of the quality aspects of both written and oral arguments. As to written arguments, MLM analyses showed that the intervention group significantly improved from pretest to posttest compared to the control group in terms of organization of arguments (F(1, 283.887) = 11.1, p = .001), with a moderate to large ES (.71); clarity of arguments (F(1, 283.928) = 4.9, p = .014), with a moderate ES (.49); elaboration of arguments (F(1, 284.041) = 9.2, p = .002), with a moderate to large ES (.65); relevance of arguments (F(1, 283.953) = 3.9, p = .024), with a moderate ES (.43); and persuasiveness of arguments (F(1, 284.012) = 3.3, p = .036), again with a moderate ES (.40).

Table 8.10 Multilevel analysis results of the quality analysis of the oral data

Measures	Fixed effects				Random effects	
	Denominator df	*F*	*P* (one-tailed)	*d*	Variance within class	Variance between class
Overall assessment	287	14.332	.000	0.93	.29	.00
Organization of arguments	279.185	22.668	.000	1.16	.23	.01
Sufficiency of arguments	280	20.567	.000	1.08	.36	.02
Clarity of arguments	278.988	26.617	.000	1.25	.35	.02
Elaboration of arguments	280.129	13.918	.000	0.91	.30	.00
Relevance of arguments	279.459	19.347	.000	1.09	.35	.01
Persuasiveness of arguments	279.283	18.595	.000	1.05	.37	.01
Addressing the opposing view	278.977	2.323	.065		.04	.00

With regard to oral arguments, MLM analyses revealed that the intervention group significantly outperformed the control group in all quality measures with a large ES except the *addressing the opposing view*[6] index, which approached significance: overall assessment ($F(1, 287) = 14.3$, $p < .001$); organization of arguments ($F(1, 279.185) = 22.7$, $p < .001$); sufficiency of arguments ($F(1, 280) = 20.6$, $p < .001$); clarity of arguments ($F(1, 278.988) = 26.6$, $p < .001$); elaboration of arguments ($F(1, 280.129) = 13.9$, $p < .001$); relevance of arguments ($F(1, 279.459) = 19.3$, $p < .001$); and persuasiveness of arguments ($F(1, 279.283) = 18.6$, $p < .001$).

8.7 Discussion

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

The developmental trends in the structural and quality aspects of written arguments were, to a large extent, similar to the development patterns in the participants' oral argument output. These developmental patterns further testify to the effectiveness of the intervention and suggest that there was a transfer of gains between the oral and written argumentative skills. The debate

[6] Note that the interrater reliability for the *addressing opposing views* variable is quite low, and hence its results should be interpreted with caution.

intervention facilitated a close connection between the arguments in the two modalities (i.e., writing and speaking) and hence made the transfer of gains easier. The debaters wrote cases to be delivered, and they presumably carried the experience they obtained after participating in each debate to the next one. In this way, the gains were strengthened and transferred between the two modalities.

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

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

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

[7] We are not aware of the use of this term in literature, which we would define as the ease with which a learner can think up arguments (i.e., a smooth flow of arguments).

Progress was also made in terms of the use of warrants as compared to the control group, with significance reached for the oral argumentative discourse. The presence of warrants indicates that the debate intervention infused in the students the ability to recognize the need to justify the link made between the data and the claim. Warrants, which are seldom employed in student argumentation (e.g., Cheng, 2010; Crammond, 1998), enhance the rhetoric and persuasiveness of an argument (Crammond, 1998).

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

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

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

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

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

8.8 Research directions and limitations

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

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

argumentation with genres other than argumentative writing (e.g., letters and reports) to gain insights into the transferability of the effects (i.e., genre-independent effects).

8.9 Conclusion

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

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

We hope that the findings of this study will encourage L2 instructors to consider employing in-class debates on a regular basis in their teaching practice. We believe that in-class debates need to become an integral part of the L2 curriculum. Debates as a teaching tool are credited and welcomed by students and hold potential for honing other skills, including L2 language skills (e.g., El Majidi et al., 2020; see Chapter 7). So they are worth the effort, as the potential payoff is substantial.

Chapter 9

Debate Pedagogy as a Conducive Environment for L2 Argumentative Essay Writing

This chapter has been submitted as: El Majidi, A., de Graaff, R., & Janssen, D. (submitted). Debate pedagogy as a conducive environment for L2 argumentative essay writing.

Abstract

Debate pedagogy may hold great potential for improving the 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 analyzed argumentative essays composed by the students on three occasions (pretest, posttest, and delayed posttest). The students' essays were analyzed 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: debate, L2 writing pedagogy, argumentative essays, L2 writing, writing development

9.1 Introduction

Many students, especially at secondary schools, struggle 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) 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) examination, 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 is heavily marginalized and under-researched (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; see Chapter 7). However, the existing evidence about the conduciveness of debate pedagogy to improving argumentative essay writing is largely anecdotal. 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 analyzed linguistic features relating to accuracy, fluency, syntactic and lexical complexity, and cohesion, as well as communicative adequacy.

9.2 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). Firstly, 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 a social and reader-oriented discourse that imparts the writer's thoughts and perspectives. Secondly, with debate pedagogy, students engage in writing (i.e., case writing) with a communicative purpose in mind, that is, defending their propositions and criticizing those of their opponents. Here, writing serves the function of synthesizing, analyzing, and organizing arguments. This orientation intersects with the WLC perspective. Lastly, the competitive debate environment stimulates the

[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.

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 marshalled 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; see Chapter 7). This self-conscious use of language seems to correspond to the WLL orientation. In brief, "debate can be seen as both a content-based and language-based subject" (Anderson, 2016, p. 74).

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 propitious for 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 the audience offers students extra impetus to develop rich, complex, and persuasive reasoning (Chen, Hand, 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., in press; see Chapter 6). 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 entail the use of higher-level reasoning strategies. Involvement in debate encourages students to critically analyze the opposing side's reasoning and evidence, and to identify inconsistencies and inadequacies in their line of reasoning. 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., 2021; see Chapter 8). The debaters in that experimental study displayed a marked tendency to diversify their arguments with sophisticated structural components. They also tended to support their positions with cogent and well-reasoned evidence.

Lastly a factor that may boost the learning process in the debate context is learners' positive attitude towards debating (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).

9.3 Debate-argumentative writing research

Very few studies (either in the L1 or L2 context) have explored the effects of in-class debate 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., Anderson, 2016; Lustigova, 2011). There are, however, two experimental studies that are noteworthy. The first study 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 employed expository essays, which she analytically assessed on the basis of three components: argumentation, organization, and communicative quality, using a 1–9 scale. The results revealed 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 are affected by debating. In another study, El Majidi et al. (2020; see Chapter 7) examined the effects of debating on free-opinion tasks. This study revealed 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).

9.4 This study

In short, several theoretical and pedagogical perspectives suggest that the debate environment can be conducive to improving argumentative essay writing. However, there is a lack of empirical research that has examined 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) for identifying lacking 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:

RQ.1 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.

9.5 Method

9.5.1 Participants

The study comprised a new sample of 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. The English proficiency level

of all classes (including writing) roughly spanned the B1 and B2 levels (the third and fourth levels of English in the Common European Framework of Reference for Languages [CEFR]; these levels are comparable to the intermediate and upper intermediate levels) as estimated by their teachers and related to Dutch educational guidelines.

9.5.2 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.

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. All debates had three phases: constructive speech, rebuttal, and clash (see Chapter 3).

Each debate consisted of three stages: pre-debate, during-debate, and post-debate (see Appendix B). Table 9.1 presents the writing activities performed in each debate session, in addition to the writing tasks undertaken in parallel by the control group.

Table 9.1 Main writing activities performed during the intervention

Intervention group	Control group
Pre-debate stage - Reading two articles and summarizing them. The instructor provided one article and the students had to find another one - Writing cases of 80 words	- Reading and summarizing newspaper articles (e.g., *the Guardian*), covering current issues, including argumentative articles, addressing, for example, political and changes in policy issues
During-debate stage - 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	- 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
Post-debate stage - Processing the feedback provided by the instructor on written cases	

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

received 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 core 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.

9.5.3 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. 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 ($n = 36$)[2]. Ethical approval for the research was granted by the researchers' institution, and consent was also obtained from the students' parents.

9.5.4 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: (a) selecting common measures used in comparable studies to ensure comparison with previous research; and (b) each measure should address a specific facet of the construct

[2] 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.

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.

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).

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 analyzer (Lu, 2010), which was specifically developed to parse L2 written data.

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 (see Yoon & Polio, 2017 for the operationalization and examples of the first three measures). 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

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: (1) average word length, which reflects mean word length, with longer words indicating more sophistication; and (2) word frequency index, which calculates the mean logarithmic frequency for all words. A lower WF indicates higher sophistication.

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

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).

According to Pallotti (2009), communicative adequacy can be measured in multiple ways, for example, by means of qualitative ratings. In this study, we used a five-point Likert rating scale that was developed in the study of El Majidi et al. (2021; see Chapter 8). This rating scale

assesses different dimensions of argumentation—which is at the heart of the communicative success of the present task—in addition to the features that ensure the presentation of an organized and structured essay. These dimensions include organization (i.e., the extent to which the arguments are well organized), sufficiency, comprehensibility (clarity), elaboration, relevance, and persuasiveness of arguments, in addition to addressing the opposing view and overall assessment.

9.5.5 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.

9.5.6 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) × 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.

9.6 Results

Table 9.2 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 9.3 illustrates these statistics for the communicative adequacy measures. Results of both Table 9.2 and 9.3 show that the groups' means increased on most measures over the intervention period and three months later.

Table 9.2 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	
Fluency	Number of words	252.30	(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)
	C/T	1.70	(.06)	1.77	(.07)	1.74	(.08)	1.87	(.06)	1.76	(.06)
Lexical complexity	MLTD	60.63	(3.00)	64.43	(3.06)	65.33	(3.57)	58.92	(2.99)	58.91	(3.01)
	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)
Accuracy	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	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	0.46	(.07)	0.33	(.07)	0.39	(.09)	0.49	(.07)	0.44	(.07)
Cohesion	Transition markers	6.92	(.46)	9.14	(.47)	8.21	(.61)	6.84	(.45)	6.70	(.46)
	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	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	5.34	(.32)	8.18	(.33)	9.01	(.43)	5.45	(.32)	6.15	(.32)

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.

Table 9.3 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)

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 9.4 and in Table 9.5 for communicative adequacy.

Table 9.4 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	MLT	285.417	5.587	.010	0.58	6.82	.67
Complexity	MLC	287.645	0.152	.348		1.22	.04
	C/T	285.871	4.249	.020	0.52	.11	.01
Lexical	MLTD	286.172	1.175	.140		197.15	23.67
Complexity	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

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

Table 9.4 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 showed a comparatively 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.

No significant differences were found for lexical complexity measures, in spite of modest observable differences in means for the MTLD 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.

Table 9.5 Multilevel analysis results for communicative adequacy

Measures	Fixed effects				Random effects	
	Denominator df	*F*	*P* (one-tailed)	*d*	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

Numerator df = 1

As Table 9.5 displays, 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[3].

9.7 Discussion

The prime 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). Results, on the whole, reveal that the intervention group significantly outperformed the control group in most of the examined dimensions of essay argumentative writing. After the intervention, the intervention 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

[3] 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.

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.

Writing in the debate context is a social activity that severs a purposeful and meaningful 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 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, Sasaki et al. (2020) concluded that writing tasks that involve a real audience can improve L2 writing fluency and motivation for writing.

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).

Debate pedagogy may also facilitate an effective interface between speaking and writing in a way that boosts writing development. In such a context, 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). This close and harmonious conjunction of language skills 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, Hand, 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.

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.

Fluency. The participants in the intervention group comparatively produced longer essays after the intervention. This is a strong indication that these participants ameliorated their argumentative writing competence. 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 have may have contributed to this significant improvement. One of these factors has to do with the positive impact of debating on argumentation competence. In an experimental study, we demonstrated that active participation in debates improved many aspects of argumentation, including argumentation quantity (El Majidi et al., 2021; see Chapter 8). 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).

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 elaborated arguments with warrants) (El Majidi et al., 2021; see Chapter 8). It appears that the intervention students needed more complex syntactic structures to formulate complex arguments.

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., Aclan & Aziz, 2015b) 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.

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 posttest, the intervention group significantly outperformed the control group after the intervention. This is consistent with the findings of El Majidi et al. (2020; see Chapter 7) 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).

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 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 establishes 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).

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). As the findings demonstrated, debate pedagogy exerted a sizable effect on all facets of this construct. This finding is in congruence with our hypothesis and in line with previous research that revealed that debate pedagogy can effectively serve the argumentative discourse, which is at the core of the communicative adequacy of argumentative essays (Oros, 2007; Zorwick & Wade, 2016). The debate environment possesses processes, incentives, and qualities that orient the mind towards nurturing a sophisticated view of argumentation and developing 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, Hand, 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, Hand, et al. (2016) contended that the audience gives students a strong impetus to develop rich, complex, and convincing arguments.

9.8 Limitations and conclusion

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

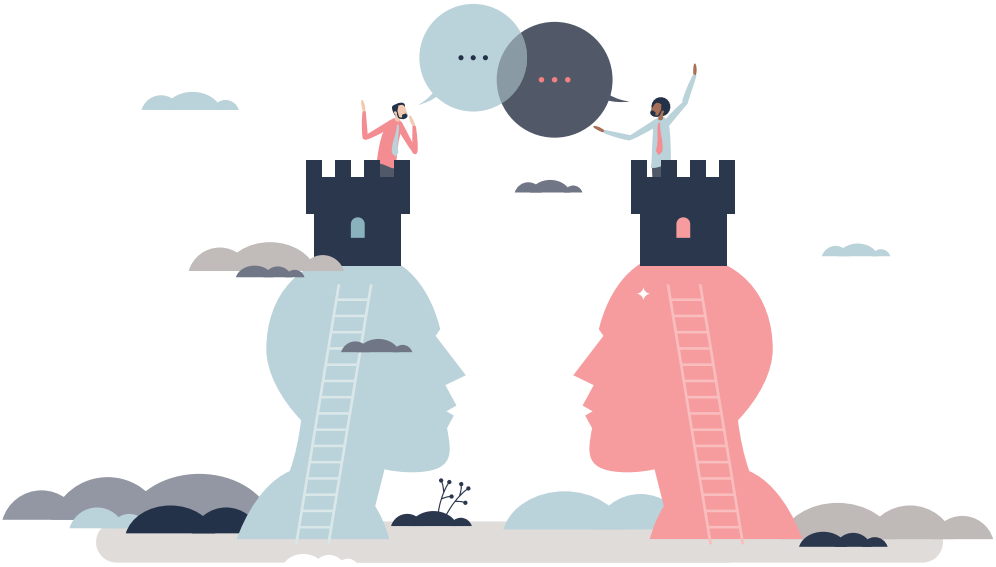
classes. Therefore, caution needs to be exercised when interpreting the long-term effects of the intervention. Our study was further limited in terms of its sample, which only included students from havo and vwo tracks (relatively high-achieving students). To get a more comprehensive picture of the intervention, the scope of future research should be broadened to include other tracks of secondary education (e.g., vocational students).

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 constitutes 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.

The present findings have important implications for L2 argumentative essay writing pedagogy. It is high time we reconsidered our conceptualization of debate as a tool that may only benefit oral and argumentation skills. As demonstrated in this study, the debate environment possesses many mechanisms and processes that are conducive to, and propitious for, argumentative writing development. 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.

Chapter 10

General Discussion and Conclusion



"The foundation of learning a second language—the four basic skills of listening, speaking, reading and writing—are all naturally part of the debating process … This is why Debate should be a core subject in an EFL curriculum. The language skills practiced and developed through debate are important skills for any second language learning. The cognitive and critical thinking skills applied through debate are also the skills required in higher learning programs." (Anderson, 2016, pp. 76, 85)

Abstract

This chapter connects key points of this dissertation. It provides an integrated overview and discussion of the outcomes of the studies in this dissertation. The chapter starts with a brief overview of the research objectives. Next, the most important findings of the studies presented in the previous chapters are summarized. In Section 10.3, I delve into these findings and weave them together to tease out the pedagogical mechanisms that may render debate pedagogy effective. In this section, I draw on supplementary data (interview with students) to further illuminate these findings. Then, implications for teaching practice are addressed followed by limitations and suggestions for future research.

10.1 Introduction

The primary aim of this dissertation was to gain insights into the effects of in-class debates on the second language (L2) speaking, writing, and argumentation skills of Dutch secondary school students. This PhD study also explored the factors that shape students' attitude towards debating and their perceptions of its effects on their language skills.

There are a number of studies that have investigated debate-related effects on language and argumentation skills, but these studies were mainly conducted in higher education and were anecdotal. This dissertation includes intervention studies that have explored the effects of debate pedagogy on different dimensions of speaking, writing, and argumentation skills, employing control groups to provide a reliable framework for understanding the contribution of this pedagogy to the performance areas being addressed.

10.2 Summary of main findings

10.2.1 Literature review of debate as a teaching tool

In Chapter 2, I conducted a literature review to fulfill two primary aims: (1) highlighting the research gaps that motivate this PhD study; and (2) identifying instructional guidance that informs the construction of a workable debate task design that forms the basis for the conducted interventions.

The literature review listed a number of benefits that are associated with debating. The review mainly focused on the improvement of language and argumentation skills because these skills are of great importance to L2 students. A number of studies have attempted to investigate the impact of debate on these skills. However, most of these studies were conducted in the first language (L1) and university context and were based on students' perceptions and instructors' observations. The studies that have investigated the impact of debate on language and argumentation skills in the secondary education L2 context are very scarce and are predominantly anecdotal. The review concludes that empirical research is needed to establish a direct causal link between debate pedagogy and the improvement of both language and argumentation skills.

As to the second aim of the literature review, it appeared that the bulk of the available instructional materials and procedures regarding the implementation of in-class debates emanated from the L1 context, which drew heavily on competitive debate. Materials and procedures tailored to the L2 context are scarce, and the existing ones provide no or few guidelines about how to stage debates in the L2 Dutch teaching practice. In addition, most of these materials and procedures were developed for universities.

Given that the main purpose of this PhD study is examining the effects of L2 in-class debates on secondary students' language production and argumentation skills, constructing a comprehensive debate task design that meshes with the Dutch teaching practice specificities is overarching. Without this task design, the potential learning outcomes of debates are unlikely to be realized and cannot be reliably and validly researched. Chapter 3 was dedicated to fulfilling this aim.

10.2.2 Debate task design

I established in Chapter 2 that the development of a balanced debate task design that explicates the mechanics of the debate activity is a key prerequisite for assessing the effectiveness of debate pedagogy. Therefore, Chapter 3 was aimed at engineering an accessible and workable debate task design that lends itself to implementation in the L2 Dutch secondary school context. Such a debate task design would eventually enable us to adequately and reliably answer the main research questions of the current PhD study.

Using the educational design research (EDR) approach, I built a debate task design and refined it through a number of iterations. In accordance with the principles of EDR, the study proceeded through three main stages. In the preliminary analysis stage, I conducted a context analysis to identify the educational problems that debate can potentially respond to. Context analysis identified three educational problems: poor L2 speaking, writing, and argumentation skills. Debate pedagogy is believed to afford ample opportunities to practice and attend to processes that help learners to enhance these skills.

After context analysis, a literature review was carried out to extract design principles to inform the development of the conceptual framework on the basis of which the initial prototype of the study was constructed. The literature review led to the extraction of 10 design principles, which were of two sorts: macro- and microdesign principles. The macrodesign principles concerned general aspects applying to the whole intervention, and the microdesign principles concerned specific aspects and activities of the intervention.

In the prototyping phase (the iterative design phase), which consisted of iterative research microcycles, the initial prototype was put to the test and subsequently modified and refined in iterative research cycles until the final prototype embodying the intended outcomes was reached. Classroom observations, group discussions, questionnaires, and interviews guided the modifications made to the debate task design. Two classes, namely *vwo 3* and *vwo 4* (the third and fourth year of the preuniversity secondary education track), were involved in the process of testing and refining prototypes. The final prototype was assessed in *havo 5* (the fifth year of higher general secondary education track).

10.2.3 Students' attitude towards debates

The study of Chapter 4 investigated the extent to which L2 students enjoyed debates and the factors that underlay their positive attitude towards this instructional tool. While there are some studies that explored these factors in L1, no study has examined the factors that generate and sustain this positive attitude in the L2 context. Therefore, this study discussed the factors that shape L2 students' attitude towards debate and the extent to which these factors differ between male and female students.

The study involved 44 students and employed a questionnaire and interviews to answer the research questions. To generate and specify the critical concepts of the questionnaire, a small-scale exploratory qualitative study informed by the existing literature was undertaken in the form of a series of focus group and one-to-one interviews. As a result, a questionnaire was developed and was piloted with a group of 18 students comparable to the target sample.

The developed questionnaire consisted of 40 items addressing various constructs (factors) that potentially shape L2 learners' favorable attitude towards debate. These constructs are *attitude*, *fun*, *active participation*, *critical thinking*, *challenge*, *teamwork*, *language proficiency* and *debate vs. course book*. In addition to the questionnaire, 10 students were interviewed to gain deeper insights into the obtained findings.

The results revealed that the participants perceived that all the constructs played a role in generating the positive attitude they have towards debate. The *course book vs. debate* factor was rated as the most influential factor with an average mean of 4.28/5 followed by the *fun* factor with an average mean of 3.98/5. The questionnaire outcomes were further confirmed by the participants in the interviews who emphasized that all the discussed factors played a role in influencing their attitude towards debate.

An independent samples *t*-test revealed that the female participants found debate more fun than the male participants. Also, the test showed that the female participants attached more importance to the *active participation* and *challenge* factors than the male participants. Interestingly, the male participants extended more importance to the *critical thinking* factor than the female participants.

10.2.4 Students' perceptions of the effects of debate

Though many studies have applauded the pedagogical qualities of L2 debate, little research has been conducted to investigate the extent to which participation in debate results in language development. One way of researching the potential effects of in-class debates on learning is through investigating students' perceptions. Hence, this study was carried out to elicit secondary L2 students' perceptions and perspectives regarding the impact of in-class debates on

different aspects of language development, namely the skill areas of speaking, writing, reading, and listening, as well as vocabulary and grammar. In addition, since research has revealed that task attitude can function as a driving force variable that may shape students' perceptions and performance (e.g., Dewaele et al., 2018), the study examined the correlations between attitude towards debate and the perceived effects as well as performance (i.e., marks).

The results revealed that the participants hold a positive attitude towards debate and correlate participation in it with an improvement in their speaking, writing, reading, and listening skills, as well as their vocabulary and grammar. The participants perceived that their vocabulary and speaking skills improved the most. During the interviews, the interviewees reinforced the facilitative role of debate in improving different aspects of the language areas being assessed and further provided examples that accounted for the effects.

The majority of the students received a positive assessment (i.e., mark). The marks ranged from 5/10 to 9.2/10. As for the students whose performance was assessed twice, a paired samples *t*-test revealed that all the debate and performance aspects being assessed significantly improved from Assessment 1 to Assessment 2.

The relationship between attitude towards debate and perceived effects was investigated using correlational analyses. There was a significant, positive correlation between attitude and perceived effects in all skills (listening, $r = .47$; reading, $r = .41$; speaking, $r = .38$; writing, $r = .32$) as well as vocabulary ($r = .47$) and grammar ($r = .38$). These correlations seem to suggest that the attitude variable tends to positively shape the students' perceptions of the effects of debate on their language development.

The relationship between attitude and debate marks was also examined. The correlation between these two variables was not significant ($r = –.06$, $p < .49$). This indicates that students of all abilities, low or high, appreciate debate. This is an interesting finding because not only do the higher-ability students seem to admire debate but apparently the lower-ability students do as well. Lastly, the relationship between debate marks and average marks for English was also investigated using correlational analysis. There was a strong, positive correlation between the two variables ($r = .58$, $p < .001$).

Chapters 6–9 present the findings of the effects of a debate intervention (developed in Chapter 3), which was conducted in three secondary schools and involved intervention and control groups, on different aspects of speaking, writing, and argumentation skills.

10.2.5 The effects of debate pedagogy on speaking skills

Chapter 6 explored the potential of debate pedagogy for developing different dimensions of speaking skills. We elicited speech samples from opinion tasks, which we coded and analyzed for measures tapping different aspects of L2 speaking competence.

The main finding of the study is that the intervention group outperformed the control group in most variables tested, with many differences reaching significance. After the intervention, the debaters produced a lengthier speech which they delivered at a faster pace and which was more accurate and coherent and lexically more sophisticated than the speech of their peers in the control group. These results generally suggest that debate pedagogy can play a facilitative role in enhancing L2 speaking proficiency.

10.2.6 The effects of debate pedagogy on written opinion tasks

Chapter 7 discusses a study that aimed to examine the potential of debate pedagogy for fostering different aspects of L2 learners' writing ability. Debate is commonly associated with oral communication and argumentation skills. Yet, debate may also offer advantages as an effective vehicle for L2 writing instruction. This study, therefore, set out to evaluate the effects of debate pedagogy on different aspects of the L2 writing competence of Dutch secondary school students.

Here, we analyzed short opinion-writing tasks (in which the students argued for or against a controversial topic) produced by the students during the pretest and posttest. 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. The students in the intervention group wrote significantly longer texts that exhibited more syntactic complexity in terms of phrasal/clausal complexity, included more sophisticated vocabulary, demonstrated better grammatical accuracy, and used a more sophisticated and wider range of indices of cohesion after the intervention than their counterparts in the control group. The study attributed the ensuing results to pedagogical features related to debate pedagogy that facilitate attending to linguistic processes that enhance L2 writing development.

10.2.7 The effects of debate pedagogy on argumentation skills

The study presented in Chapter 8 examined the effects of L2 debate pedagogy on argumentation skills. It investigated the extent to which this pedagogy affects a number of L2 structural (e.g., subarguments and rebuttals) and quality (e.g., elaboration and persuasiveness of arguments) aspects of both written and oral argumentation.

To answer the research questions of the study, we analyzed written and oral opinion tasks produced by the participants at the beginning and towards the end of the intervention, employing an adjusted Toulmin model of argumentation to undertake the structural analysis and a five-point scale rubric to assess different aspects of reasoning quality.

The findings, on the whole, showed that debate pedagogy positively affected a number of structural and quality aspects of the argumentative competence of the intervention participants'

written and oral argumentation. After the intervention, the debaters displayed a marked tendency to diversify their arguments with sophisticated structural components and to utilize good-quality reasoning. The developmental patterns in the structural and quality aspects of written and oral arguments were largely the same.

10.2.8 The effects of debate pedagogy on argumentative essay writing

Chapter 9 reports the findings of the study that investigated the potential of debate pedagogy for improving the quality of argumentative essays. Chapter 7 investigated the impact of debate pedagogy on L2 writing performance and Chapter 8 focused on the argumentation aspects of the produced texts. However, these texts were relatively short, and hence those findings cannot be conclusively taken to indicate that debaters can compose good longer structured argumentative texts (i.e., argumentative essays). Therefore, the study of this chapter was set up to examine the extent to which debate pedagogy can affect different areas of argumentative essays and to identify the scope of the effect.

To measure the effects of the intervention, we analyzed argumentative essays composed by the participants on three occasions (pretest, posttest, and delayed posttest) for different textual and content features. Results, on the whole, showed that the intervention group significantly outperformed the control group in most of the examined dimensions of argumentative essay writing. After the intervention and during the delayed posttest, the intervention group wrote argumentative essays that were longer, exhibited more complex syntactic structures, and were formulated in a more diverse, coherent, and persuasive language.

10.3 Discussion of main findings

In the summary of the main findings above I have presented the results separately. I will now weave these findings together and discuss the pedagogical attributes that presumably underlie the obtained positive learning outcomes. These attributes are distilled from the interpretation of the results emerging from the conducted studies. This research, to the best of my knowledge, is the first investigation to have empirically examined the impact of debating on different dimensions of speaking, writing, and argumentation skills in the L2 secondary school context. The act of speaking, writing, and reasoning in this environment promotes language and argumentation performance in ways that have not been fully acknowledged.

This study has provided research-based evidence on the affordances of debate pedagogy. The findings, on the whole, suggest that well-structured and well-run debates may help learners to enhance their oral, written, and argumentation performance. This departs from the common view that debates are only apt for extracurricular activities or in the competitive sphere.

The cyclic nature of debate pedagogy involves students in a meaningful environment that stimulates linguistic and reasoning processes that lead to sophisticated thinking (reasoning) and language usage. The debate intervention has facilitated piling up and reinforcing learning gains. Each debate cycle can be seen as a brick added to a foundation. As each debate cycle is completed, new learning gains are secured, strengthening the foundation. With succeeding debates extending the work achieved in previous debate cycles, learning gains are further developed and transferred to a subsequent cycle, culminating in building strong speaking, writing, and argumentative skills. In other words, I presume that the intervention has created a snowball effect in which the students developed and appropriated speaking, writing, and argumentative skills over time from different debate cycles. Generic ties between these cycles presumably facilitated the transfer of gains from one cycle to the next (see Figure 10.1).

Arguably, the cyclic learning process of debate and the coherence of the debate task components may have increased the automaticity of language production. For example, the debaters used language forms in cases[1] (which they came across in readings) and later on multiple occasions during rebuttals and clashes. This repeated retrieval of forms may have led to automatization (see DeKeyser, 2007b) and "increases the likelihood that [these forms] will be produced later in a more spontaneous setting such as conversation" (Williams, 2008, p. 13). Such an environment presumably frees up demands on working memory in favor of paying attention to more challenging aspects of production.

The debate intervention offered repeated opportunities for students to engage in a learning environment involving critical reading, writing, speaking, and listening as well as demanding outperforming critical opponents. This context requires debaters to ponder over the substance of their contributions (i.e., arguments) and on how to formulate them. As such, it stimulates them to reflect on and critique their learning by reviewing their performance. In such an environment, the very act of speaking, writing, and arguing promotes linguistic and rhetorical processing that potentially leads to language and argumentation development (i.e., writing-to-learn, speaking-to-learn, arguing-to-learn). Involvement in this learning experience empowers debaters to identify shortcomings in linguistic and reasoning skills and develop metacognitive knowledge of what constitutes accurate, sophisticated, and adequate (functional) language. Framed differently, in such an environment, debaters are "pushed toward the delivery of a message that is not only conveyed, but that is conveyed precisely, coherently, and appropriately" (Swain, 1985, p. 249).

[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.

10.3.1 Debate pedagogical profile

Debate pedagogy provides fertile ground for an interplay of many pedagogical approaches and practices that mutually strengthen and interact with each other and ultimately lead to enhanced language and argumentation performance. In what follows, I will list some of the debate-related pedagogical factors that may have led to the results that have consistently emerged throughout the studies we have conducted. The list is by no means exhaustive, and I equally believe that the observed gains hinged on the synergetic contribution and collaboration of a number of these factors, if not all of them. These features may also provide insights into pedagogical mechanisms that scaffold processes that lead to L2 improvement.

In this section, I will draw on interviews (one-on-one and group interviews) conducted with 44 volunteered students after participating in the intervention studies presented in Chapters 6–9. These interviews mainly explored the students' perceptions regarding the impact of debate on language and argumentation skills in addition to grammar and vocabulary. This interview is akin to the one conducted in Chapter 5 with the exception of one additional question eliciting the students' perceptions of the effects of debate on argumentation skills. On the whole, the students' perceptions in these interviews are congruent with the ones presented in Chapter 5.

Competitive pedagogy

During debates, students tend to push each other to the limit. They challenge each other not only in terms of the strength of the arguments they advance but also in the way they frame them. That is, their challenge does not stop at the argument strength level, but it extends to the language use level. The urge to outperform competitors seems to put extra demands on debaters for more precise, accurate, and sophisticated linguistic production both during preparation (e.g., composing cases) and during actual debates. In other words, the debate context seems to instill in students the awareness of the importance of attending to the language they employ to fulfill their objectives. Debaters are aware of the fact that they are on a mission of defending their standpoints and outshining 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 language and argumentation production as these debaters pointed out:

> [You use] longer sentences, also more effective sentences just constructing better arguments. Debate gives you a kind of motivation to beat the ones you are debating against, and it really gives you motivation. And you want to win even though it is not a real contest, but

> you know you just want to win. ... You're just going to do your best so you just use longer and more difficult sentences. (Student 1)

> If you want to come up with something good, you have to pay attention to many things, which is why I made fewer mistakes, or almost no mistakes in the last cases. (Student 2)

> You want to win from others ... so I want to have the best arguments ... if you use easy words then your arguments are not so powerful. (Student 3)

The inherent competitive atmosphere of debate fuels students to generate rich and lengthy negotiations that sharpen up their reasoning and linguistic abilities. As shown in Chapter 2, *competitiveness* is a driving motive that energizes students to actively participate in debates. Similar results emerged from previous research. For example, Gelinck (2000) reported that his participants appreciated the competitive element of debate, which stimulated them to try their best. Similarly, Cinganotto (2019) noted that debaters seem to "appreciate [the] peculiar mix of learning and competitiveness" (p. 124).

Motivating pedagogy

The effect of any pedagogical tool on learning hinges on the willingness of the learner to actively launch into task performance, take part in its interactions, and mobilize their cognitive resources to process information (Van den Branden, 2016). A number of studies have demonstrated that a positive task attitude substantially impacts learners' willingness to participate in the learning process; it functions as a filter (Dörnyei, 2002; Dörnyei & Kormos, 2000) and boosts students' motivation to perform to their utmost (MacIntyre et al., 2001; Masgoret & Gardner, 2003).

No wonder that when a learning tool is appreciated and perceived as fun and interesting, it will most likely be experienced as motivating. In other words, "the more that students believe in the value of what they're being asked to do, the more motivated they will be to pursue the tasks assigned to them" (Hirvela, 2016, p. 96). This seems to apply to debate as one student explained:

> You just like it [debate] so you want to do it ... If something is fun, then you want to start with it quickly. (Student 4)

More importantly, the motivating environment of debate seems to enhance the impact on the learning gains. Student 4, who was talking about the effectiveness of feedback in the debate environment, makes a case for this point:

> What has improved during debates stays with me longer because I am motivated. So I know exactly what I have written, and when I know, oh, that I have done something wrong, I remember it for the next time. (Student 4)

As demonstrated in Chapters 4 and 5, the students revealed that they hold a positive attitude towards debates, characterize them as motivating and rewarding, and advocate a systematic debate instruction. The interviewees further indicated that they admire debates because they see them as a viable vehicle towards attaining a number of objectives. For example, they offer them a platform for expressing opinions, improving language skills, and priming them for exams (see also Chapter 4):

> I like to write cases because I can put my opinion in them. (Student 3)

> It's a pity that debates have stopped. I wish they lasted longer. They were fun and everyone just felt comfortable, and you learn a lot from others as well as from the articles you read. (Student 4)

> I find [debate] important because I saw that it helps with exams ... debate has components that we get at the exam like reading comprehension. (Student 5)

Audience-centered pedagogy

Debating is a social learning enterprise. The social dimension is triggered by the presence of an authentic audience, which is a key element in debate pedagogy. As we saw in the studies presented in this dissertation, the audience in debate serves an important function. It seems to raise awareness of the utility of accurate and sophisticated rhetoric and language use. When students present or write something for a real audience, they become motivated to reflect upon their production (Chen, 2019).

What further distinguishes an audience in the debate context is that it also consists of critical and competitive opponents. This seemingly intensifies the functionality of sophisticated language and argumentation. These findings are echoed in previous research. For example, Chen and Brown (2012) argued that 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).

Interestingly, an audience serves an additional function in debate. The debaters reported taking pride in the strength of their arguments and the progress they have booked, and the audience, here, is needed to witness these achievements as these two students illustrated:

> If you have good arguments, you just can't wait to show them to those who are on the opposite side. (Student 6)

> You want to show your classmates how well your language skills have improved, and then you prepare well. (Student 7)

Engaging pedagogy

Active involvement of students in the learning process is one of the secrets of a successful learning experience. Meyers and Jones (1993) conceptualized active learning as anything that "provides opportunities for students to talk and listen, read, write, and reflect as they approach course content" (p. xi). Bonwell and Eison (1991) pointed out that students learn actively when processing content through analyzing, synthesizing, evaluating, discussing, and using it in meaningful ways. Omelicheva and Avdeyeva (2008) pointed out that active engagement in the learning process can be further boosted through engaging students' emotions, attitudes, feelings, preferences, and values. To achieve this, learners need to be afforded ample opportunities to communicate verbally (Bellon, 2000).

Darby (2007) contended that debate possesses a rich profile of active involvement, as it actively involves students in a rich learning process that induces them "to research critical issues, prepare and present a logical argument, actively listen to various perspectives, differentiate between subjective and objective information, ask cogent questions, [and] integrate relevant information" (p. 1). MacArthur et al. (2002) observed their debating participants continued to debate beyond the allowed time limit, and they stopped only when the teacher intervened. This comment concurs with my observation that in many debates the debaters carried on with debating even after the conclusion of the debate sessions.

The interviewees acknowledged the role of debate in stimulating them to take on an active role during debate. For example, the following students maintained that:

> I liked them [debates] because everyone is involved … because it [debating] is active, you have to participate. To be able to speak, you need to listen well. When someone says something, you cannot respond if you are not listening carefully. (Student 2)

> I find them instructive because you are actively involved [in the learning process]. You are not only active in [debate] lessons but also at home [because of preparation]. (Student 8)

Interactive pedagogy

Debate is a multidimensional interactive pedagogical tool. The interaction manifests itself at different levels (see, for example, Chapter 5). Wade (1998) lauded the pedagogical interactive

merit of debate as he stated that "there are certainly trends in education which encourage interactive and dialogic pedagogies, but few are as potent as debate" (p. 63). Debate-prompted interactions avail learners of an array of learning opportunities and processes that cater to language learning. This student alluded to this when he said that:

> It [debate] was quite informative ... When you really talk with each other, I think you develop your English more than when you learn from the books. (Student 9)

The different interaction patterns facilitated by debate create opportunities in which, and through which, learning occurs. During these interactions, debaters challenge each other's line of reasoning, and this competitive context fuels interactions and induces long and complex turns and use of discourse strategies (see Nakahama et al., 2001). In other words, these interactions enable debaters to jointly construct and scaffold each other's learning processes, as they provide each other with input and create opportunities for pushed output as this student pointed out:

> Speaking skills improve because you are pushed to talk… You have to talk and as a result you learn new words and everything. And when you hear others talk, you learn new things. You also hear the pronunciation of others. You can also learn from this. (Student 10)

Previous research made a case for the facilitative role of negotiated social interaction in promoting language learning. It can operate "as a source of input and opportunities for output which foster the internal processing that results in acquisition" (Ellis & Shintani, 2014, p. 194). Output prompts a cognitive comparison between input and learners' linguistic problems and accordingly results in improved performance (e.g., Izumi et al., 1999). Interactions also afford learners opportunities to gain extra control over existing linguistic resources (Ellis & Shintani, 2014). In addition, dialogic social interactions provide learners with insights into their reasoning gaps. Noticing and hence becoming aware of these gaps allows debaters to refine and develop sophisticated thinking.

Collaborative pedagogy

Research has shown that students talk more and use a wider range of speech acts in the group context, especially in small groups (Ellis & Shintani, 2014; Long & Porter, 1985). Group work provides learners with multiple opportunities to use L2, and it promotes learner independence (Ellis & Shintani, 2014). Li et al. (2020) and Varonis and Gass (1985) also credited the effectiveness of L2 conversations as a nonthreatening context in which learners can practice language skills and make input comprehensible through negotiation.

Roy and Macchiette (2005) maintained that debate promotes a sense of teamwork. In debate, "the students have to cooperate and help each other find the best resources and the most convincing evidence" (Cinganotto, 2019, p, 115). Zare and Othman (2015) revealed that learners appreciate the nature of the collaborative work being required by debates. In the same vein, Fallahi and Haney (2007) reported that "a total of 80% [of their participants] experienced a feeling of group accomplishment or teamwork during the debate" (p. 86). Students in this study echoed these findings, for example, when they were juxtaposing a one-to-one with a two-against-two debate format that allows collaboration:

> Two against two is more fun than one against one because you are not alone. You do not have to defend yourself alone because someone else has the same point of view. Someone is sitting next to you. They can help you. (Student 11)

> Everyone experiences two against two as better and more fun because you work together. And if you don't know something, your partner can help you. (Student 12)

Student-centered pedagogy

With student-centered pedagogy students are placed at the center of teaching and learning, and the teacher takes the role of a coach, advisor, and facilitator. Altinyelken (2015) pointed out that student-centered pedagogy holds potential for promoting active learning, stimulating the use of multiple teaching and learning methodologies, and enhancing classroom participation through increased interactions among students, as well as between students and teachers. Weinberg (2017) contended that student-centered learning stimulates learner choice, which can provide a sense of ownership over conducted tasks and enhances self-determination and enjoyment. Johnson (2003) stated that a student-centered pedagogy is intrinsically motivating for L2 learners because it takes into account their learning needs and styles and allows them to have a say in the kind of activities they prefer.

Debates lend themselves readily to the philosophy of student-centered pedagogy. Well-designed debates grant students the tools and power to manage the learning process with minimal interventions on the part of the instructor. Debates transform teachers into coaches, a perspective that stimulates more mentoring and individualized in-class help and less dominating classroom styles (Bellon, 2000). Musselman (2004) presented a detailed account of the new role debates forced her to take:

> I watched my classroom of 20 undergraduates conduct a 75-minute, critical discussion of primary sources, and I never said a word. This, I thought, was one of those magical moments for which I had always hoped: the day I would make myself virtually redundant

> as a participant in class discussion because I had given my students the tools to carry on the conversation without me. (p. 335)

I totally concur with Musselman's account. My intervention in the debate sessions was minimal and further decreased as my students became more familiar with the debate structure and mechanics. The students appreciated this pedagogical aspect of debate as can be inferred from the comment of this student:

> With debate you are active. If you only listen to teachers, some students just literally fall asleep. (Student 13)

Integrative pedagogy

Learners' output may benefit from being engaged in tasks that integrate skills and facilitate a meaningful link between input and output. Debate lends itself easily to skill integration (Alasmari & Ahmed, 2013; Lieb, 2007; Rybold, 2006; Snider & Schnurer, 2006; Stewart, 2003; Zare & Othman, 2013) because "the four basic skills of listening, speaking, reading and writing are all naturally part of the debating process" (Anderson, 2016, p. 79). The students have also acknowledged the integrative power of debate:

> You are working on different areas, reading, writing, and listening to other students. I think it would be effective for other modern foreign languages. (Student 8)

> I find it [debate] instructive. I have learned many things, learn to speak well, read, write well, and listen, so very instructive. (Student 14)

Prior to actual debates, students do some preparation during which they research the topic of debate, namely through reading critically to identify the arguments that would swing the balance in their favor. A proper preparation also entails preparing a case that encapsulates the arguments that defend their point of view and the arguments the opposing team would likely use. In addition, a meaningful debate cannot take place without students critically listening to each other's arguments to comprehend them and spot inadequacies in them. It goes without saying that the speaking skill receives the lion's share of attention in the debate process.

The harmonious and smooth integration of skills facilitated by debate pedagogy seems to enhance the effect on performance as skills are integrated "in such a way that they supported and enhanced one another" (Lustigova, 2011, p. 26). The integrative pedagogy of debate also seems to benefit content (i.e., argumentation). Anderson (2016) remarked that "it is hard to imagine a more harmonious integration of content and language skills than in the teaching of

debate" (p. 76). In what follows I will cast light on a number of skill integration configurations that figure in debate pedagogy.

Writing-to-speak. Writing can support speaking in many ways. Writing by its nature possesses many unique features facilitative of L2 language learning and development (Williams, 2012). First, because of its low pace and the permanence of its record (Williams, 2012), writing prompts learners to attend to linguistic forms and stimulates them to try new forms of language (Adams & Ross-Feldman, 2008; Belcher & Hirvela, 2008; Niu, 2009). Second, because of its cyclical nature (Kormos, 2014), writing promotes closer attention to linguistic elements (Manchón & Roca de Larios, 2007), and this leads to conscious reflection on language use and noticing gaps in linguistic knowledge (Kormos, 2014). Additionally, the writing environment gives confidence to writers about recently tried and acquired structures and forms and more importantly boosts the transfer of these gains from planned language to spontaneous production (in speaking) (Adams & Ross-Feldman, 2008; Belcher & Hirvela, 2008). In this way, oral proficiency can largely benefit from integrating written language with oral production.

I believe that the debate environment can effectively connect spoken and written discourse by creating opportunities for the two modalities to mutually support each other. In debates, writing scripts oral performance as debaters prepare cases that they turn to during the first stage of debates (constructive speech). The writing-speaking connection also manifests itself during the second (rebuttals) and third (clash) stages. Prior to these stages students are allowed to pre-plan (through making notes) their performance. Earlier research (e.g., Yuan & Ellis, 2003) demonstrated that written pre-planning empowers learners to use new and more complex forms in their oral production. The students also felt the benefits of writing-speaking connection in debate as indicated by this student:

> You should first write down what you are going to say, and you learn something from this. You remember words. (Student 15)

Speaking-to-write. On the other hand, debate pedagogy also offers channels for a speaking-writing connection. During the interactions of debate contests, debaters become aware of their linguistic gaps and even learn new words from each other (see Chapter 5). Students in the interviews reported that they used many of these words in subsequent cases and writing tests:

> I would not have been able to perform well in writing tests, for example, if I did not have them [debates] beforehand. I already knew a few words by heart that I used in debates. (Student 14)

> I used the difficult words that I found in articles and learned from classmates in cases. (Student 16)

Students' perceptions are supported by previous research that revealed that spoken interactions between learners provide scaffolding for their writing development (Hyland, 2008; Yang, 2008). Through collaborative dialogue through which students express ideas and reflect on what they and others say, learning takes place (Swain, 2000).

On the whole, the speaking and writing modalities are closely connected in debate pedagogy. In this environment, we can presume that some learning gains transfer from writing to speaking and from speaking to writing. In other words, the two modalities can "synergistically contribute to each other and to L2 proficiency overall" (Belcher & Hirvela, 2008, p. 2). This means that the development of L2 speaking and writing proficiency is bidirectional (Williams, 2008). Put in Rubin and Kang's words (2008), in a context like debate, "the arrows between development in speech and in writing are double-headed" (p. 211).

Reading-to-write. The debate pedagogical profile also nests the reading-to-write pedagogy. This pedagogical formula has proved to be beneficial to many dimensions of language and rhetorical development (e.g., Hirvela, 2016; Hyland, 2019). Each debate involves reading source texts followed by writing cases in preparation for debate contests. The debate environment seems to augment the value of reading as Oros (2007) suggested: "[S]tatistical analysis of course evaluations indicates that the use of ... debates lead to increases in student value placed on course readings and assignments when compared to courses where such debates were not used" (pp. 294–295). Ferris (2011) briefly summarized the linguistic benefits accruing from this experience by stating that "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). When L2 learners navigate source texts, they tend to extract structures and forms and transform them into what they are writing (Hirvela, 2016). This means that these source texts function as a template that helps learners scaffold different dimensions of their writing ability as these interviewees pointed out:

> Sometimes you make longer sentences because you read articles where they use longer sentences, and you actually try to copy these constructions to your style. (Student 17)

> Because I read a lot more [debate readings], I see new constructions that strike me and I use them too. (Student 18)

Importantly, as the following interviewee holds, sophisticated constructions are processed and reformulated before being used:

> You do not literally copy difficult sentences from articles, but you process them in your own words in cases. This of course improves your writing skills. (Student 6)

In addition to fostering linguistic competence, the reading-to-write pedagogy also seems to promote argumentation competence. Searching for arguments to build one's case provides debaters with "an authentic reason for them to read" (Mirra et al., 2016, p. 11). As we saw in Chapter 6, the vitality of reading-writing pedagogical formula is further enriched by adding speaking to it, or rather reading-to-write-to-speak.

Figure 10.1 provides a synopsis of the possible confluence and synergy of the pedagogical formula of debate pedagogy.

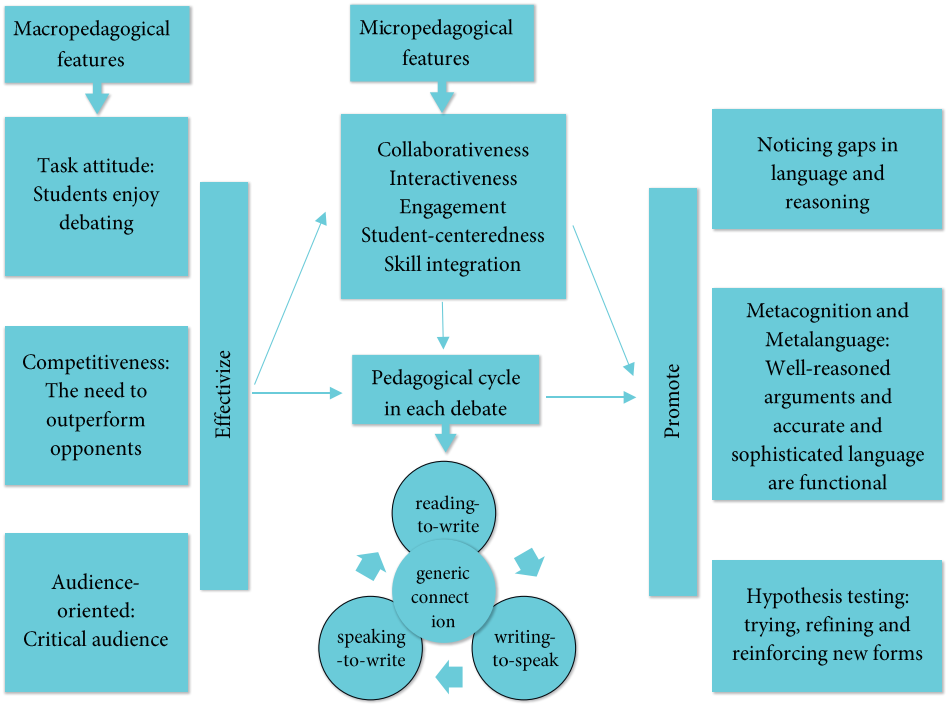


Figure 10.1 Pedagogical configuration of debate pedagogy

As such, I believe that these pedagogical qualities jointly enhance the effect on learning processes in debate. I assume that competitiveness, positive attitude, and the presence of an audience features (which I termed macropedagogical features) affect the general environment of debate including the micropedagogical features (particular pedagogical features of debate,

which include collaborativeness, interactiveness, engagement, student-centeredness and skill integration). The microfeatures, in turn, together with the macrofeatures affect the pedagogical cycle of each debate. Ultimately, this whole operation affects processes that lead to language and argumentation development. Cinganotto (2019) hinted at the contribution of some of these features to an enhanced effect on learning in the debate environment by stating: "[D]ebating in a foreign language is worth trying and experimenting…the mix of research, discussion, engagement, competition, cooperation and fun seems to have a strong potential in a language learning environment" (p. 120).

10.3.2 Debate environment as fertile ground for improving lexical competence

The study has shown that debate pedagogy facilitates the enrichment of the lexical competence of debaters. The findings in the last five chapters clearly point in this direction. Improvement of lexical competence is one of the key factors that has arguably boosted performance in other skills as these interviewees reported:

> Writing is easier [now] because of that knowledge of those words and how to use them. (Student 19)

> Because my vocabulary is enlarged, I understand texts better. (Student 20)

> I can come up with things faster now because I know more words in English and hence things go much more quickly. (Student 21)

The debate environment possesses unique features that induce the development of a great, diverse, rich, and active vocabulary base. One of these features is that read, written, and delivered texts share the same genre (i.e., argumentative genre), and research has demonstrated that when there is a close generic connection between the modalities of a task (e.g., Hirvela, 2016), words can easily transfer from one modality to another. Debate cycles and series have further contributed to the retention of these words. In addition, a strong vocabulary base is functional in the debate context; it serves the function of constructing strong arguments and increasing their persuasiveness.

Previous research also highlighted this point. For example, Mirra et al. (2016) remarked that "Bobby [a debater] felt motivated to learn new words in order to deliver powerful speeches during his debate rounds" (p. 11). These new gains (i.e., words) provide ready-made building blocks for oral and writing production. Since the debate environment offered many opportunities to implement these new words (during case writing and sometimes in other debates), their retrieval has become cognitively relatively undemanding (see Section 10.3.4).

During the interviews (as in Chapter 5), the students massively and unequivocally hailed debate for the facilitative role it has played in this regard. They particularly pointed to two sources through which words percolated into their lexical repertoire: preparatory source texts and classmates. More importantly, they underlined that they actively employed these words in one of the cycles of debates or in other contexts:

> [I learned words] especially through articles because there are difficult words in them, and I used them in debates. (Student 22)

> During preparation, you had to read articles, and I learned new words because there were many words I didn't know … and I later used them… [In addition] I came across words in the third debate which I learned from the articles of the first debate, and in this way, I learned words. (Student 15)

> A few classmates are very good at English, and during debates I learn words from them. (Student 18)

As a result of building a strong base of active vocabulary, the debaters developed a degree of fluidity and automaticity over their language production. Many debaters in the interview pointed out that after the debate intervention, they gained fast and automatic access to their linguistic knowledge (i.e., improvement of fluency):

> I can get fast to words [and this helped] my fluency. (Student 5)

> If someone asks me something on the street, I can give an answer and make a sentence more quickly than before. Before debate [intervention] I couldn't do this. (Student 20)

> I need to think less to retrieve words during rebuttal and clash …I could come up with arguments without preparation. (Student 23)

10.3.3 Debate environment as fertile ground for improving argumentation skills

The findings of the debate-argumentation study (Chapter 8) demonstrated that debates offer not only a language nurturing environment but also a supportive environment for developing argumentation skills. This means that the benefits transcend the linguistic features and extend to content or "the thinking that underlines the writing" (Kuhn et al., 2016, p. 145) and speaking. The students massively recognized that their argumentation skills had improved after the intervention. Many, for example, noticed that their argumentation fluency substantially improved:

> I can come up with ideas more quickly because I have had debate all the time. I can find arguments more quickly. (Student 5)

> In the first essay I didn't know what to think of. Because of debate I can think of arguments faster. (Student 17)

The interviewees reported that debating improved both their oral and written argumentation, and more importantly as a result they were able to produce good argumentative essays and perform well in oral exams:

> I've noticed that in the debate lessons you have to come up with a lot of arguments to support a proposition and as a result you improve your argumentation and you write better essays. (Student 4)

> I failed last year. During the speaking exam, I repeated the same argument over and over again during the discussion part, and with debating I had a different argument every time, strong arguments, and not repeating the same one every time. (Student 24)

The debate environment seems to possess many incentives that lead to the improvement of the structural complexity and quality aspects of L2 argumentation. One of these incentives is its competitive atmosphere that prompts debaters to have a critical look at their arguments in an attempt to marshal the ones that would help them outsmart their opponents. These opponents are critical; outperforming them necessitates making recourse to well-crafted and well-reasoned arguments. In the words of Kuhn et al. (2016):

> students become interested in the quality of their argumentation as a goal in itself. They want … to produce strong arguments, ones they anticipate their opponents won't be able to easily "shoot down." They strive to construct strong counterarguments and rebuttals because they appreciate them as better than weak ones (p. 25).

Some students confirmed this viewpoint:

> During preparation I pay close attention to arguments ... think about how I can ensure that the other side cannot come up with a counterargument or this argument. (Student 25)

> When you prepare a case, you think like if the other party can refute them [arguments] quickly, then you think of more strong arguments. (Student 26)

This context has seemingly developed the debaters' awareness of strong/weak arguments and more importantly nurtured their insights into how to make arguments sound and solid:

> I think that it has now become easier to distinguish the strong arguments from the poorer ones. I can substantiate those good arguments better, and when they refute them, you can quickly come up with counterarguments. (Student 27)

> You learn that these [arguments] are strong and these are perhaps easier to counter, and then you just write down the strong ones. (Student 28)

> You use more facts because they come over more strongly. (Student 29)

10.3.4 Transfer of learning and gains

The transfer of learning that involves the application of previously gained knowledge and skills to novel situations "has long been proclaimed one of education's central aims" (Green, 2015, p. 1). As Bransford and Schwartz (1999) pointed out, "transfer lies at the heart of our educational system. Most educators want learning activities to have positive effects that extend beyond the exact conditions of initial learning" (p. 61). The transfer of learning is key to determining the effectiveness of what we teach. It is an aspiration that every instructor entertains, hoping that their students will apply what they have been taught "whether in a future class session, another course, or an area of life beyond school" (Hirvela, 2016, p. 87).

Debate seems to promote a positive and productive environment for transfer. There are different relevant transfer manifestations and dimensions that the debate intervention has inspired, ranging from transfer of gains between skills within and beyond the context of debate to transfers to other subjects and beyond the school context. Though all transfer patterns reported by the students are of paramount importance and testify to the effectiveness of debate pedagogy, the transfer of gains to novel contexts unrelated to the realm of debate might even constitute a stronger proof that strongly accounts for the transferability of the obtained effects.

Transfer of learning gains between skills is arguably facilitated and prompted by debate pedagogy, which seems to cement the connection of skills in different pedagogical configurations, such as reading-to-write and writing-to-speak. Transfer of learning did not go unnoticed by the debaters. They felt and recognized the transportation of learning between skills:

> I had to read articles. I came across a lot of words I didn't know. I translated them and I started using them in cases and after cases also in my rebuttal, and I used them in other debates. (Student 15)

> I use them [connecting words] more often even beyond the debate context. (Student 30)

The interviewees also reported that the gains from the debate experience transferred to contexts in which their L2 performance was assessed:

> During the writing exam (letter), I used words I learned in debates. (Student 22)

> At the beginning of the year, I had sevens for writing (letters), and at the end of the year I had 9.5/10. I think that's because of debate. (Student 31)

As to the effects of debating on argumentation skills, the debaters felt that this effect transferred to other subjects:

> Debate helps to think critically and listen critically to your opponents and to come up with good arguments in the clash, and critical thinking has helped in other subjects where you can apply it. (Student 20)

> Also, in Dutch [subject] my argumentation skills are better now. (Student 32)

Lastly and most importantly, the debaters pointed out some examples of transfer beyond school:

> I have realized that I can communicate more quickly with relatives from England. (Student 1)

> I come across people who don't speak Dutch. In the beginning I would say how do you say that and then realized that, oh, I had this during debate. I can say so and so. I can now find words quickly. (Student 33)

These accounts of transfer-of-learning again suggest that the debate environment possesses mechanisms that stimulate a productive transfer climate. The students seem to contend that they managed to retain gains they obtained from the cycles of the debate experience and more importantly to transport them into new contexts.

10.4 Implications and recommendations for teaching practice

This dissertation has shown that debates are motivational and allow students to engage in a meaningful learning experience. The debate learning experience is meaningful because debate 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 in both oral and written contributions. Debate pedagogy also creates an interactive and rigorous learning environment that facilitates a synergetic integration of language skills that mutually support and reinforce each other. In this environment learners are exposed to relevant and meaningful L2 input and are prompted to produce sophisticated, accurate, and adequate L2 output. As shown in Section 10.3.1, debate pedagogy seems to nest a range of unique combinations of pedagogical features

that presumably activate processes that lead to the improvement of many dimensions of speaking, writing, and argumentation skills.

In-class debates provide unique learning opportunities and hence warrant inclusion in L2 curricula. While debating, students are doing something worthwhile that leads to substantial progress. This has been recognized by some students who even recommend debates for other foreign languages:

> You are busy with different areas, reading, writing, and listening to other students. I think it [debate] would be effective for other modern foreign languages. (Student 8)

> Look at French, for example, at the end [of secondary school studies] you get your diploma but no one speaks French. Through debating in the French, German [classes] or other languages, you learn more, also to speak [these languages]. (Student 9)

As the interviewees suggest, debate should be an integral part of not only teaching English but of other foreign languages as well. Successful and effective debate instruction could benefit from a constructive collaboration between language departments at schools including the Dutch language department. They can set up a debate task design that fits their teaching perspective and context. The proposed debate task design (in Chapter 3) lends itself readily to adjustment to multiple environments even beyond the Dutch educational context. Importantly, I also believe other school subjects (e.g., history, philosophy, and social studies) can benefit from employing debates. As mentioned in Chapter 3, debates can lead to a deeper understanding of course content. It is true that implementing debates in class often places extra demands on the instructor, but they are worth the extra time and effort they require.

The pedagogical attributes of debate have provided us with insights into mechanisms that render debate and expectedly other teaching tools effective. As such, the study makes a potentially valuable contribution to L2 pedagogy by highlighting a number of pedagogical practices that can be advantageous to enhancing L2 learning. The study further suggests that when these pedagogical practices are implemented in an engaging and meaningful environment, learners are likely to mobilize their efforts and afford the learning process their full attention. In short, the pedagogical profile of debate can act as a worthwhile guiding framework that can inform the selection of effective tasks and how to augment their effectiveness.

Instructors need to occasionally have the confidence and courage to put course books aside and choose contents and pedagogical tools that interest learners because this decision would result in increased involvement and motivation and hence boost performance. Tasks that are meaningful, challenging, interactive, student-centered, and audience-oriented, and that cater to students' interests and involve their emotions and opinions, are likely to be motivational and

instructive. In-class debates are not meant to supplant course books or other methods of instruction, but they can be used to supplement them.

Textbook publishers also need to recognize the potential of debates and systematically include them in their textbooks. Importantly, they need to provide instructors with a clear and flexible roadmap of how to stage debates in practice from start to finish.

10.5 Limitations and suggestions for future research

This PhD study has a number of noteworthy limitations that open up future research avenues that may further unravel the pedagogical potential of this tool. Most of these limitations have already been discussed in the conducted studies in each chapter. Therefore, I will briefly touch upon these limitations, which mainly concern five areas. First, when tracking the effects of debate pedagogy (on speaking, writing, and argumentation skills), we only analyzed one task for each participant. One task is unable to conclusively capture a complete picture of the tracked effects. To obtain full and fine-grained insights into these effects, future research needs to elicit and analyze more data per participant. Second, with the exception of the last study (Chapter 9), other studies lacked delayed posttests, which could have provided relevant insights into the long-term effects of the conducted interventions. Third, this study assessed students' performance in tasks that may have fairly favored the intervention students over their control counterparts. Since performance tends to vary across task types and genres, it is important to employ a variety of different tasks to gain a deeper understanding of the effects of debate pedagogy. Fourth, our sample is not representative of all secondary school students, as it only includes havo and vwo students (who are relatively higher-achieving students). It would be worthwhile undertaking this research with lower-achieving students (*vmbo* students, prevocational students) to further deepen our understanding of the potential of debate-based instruction. Last but not least, this study endeavored to report the impact of debate pedagogy in three Dutch schools, which may not be representative of all Dutch secondary school students, let alone secondary school students beyond the Dutch context. Therefore, it behooves us to conduct this research in other contexts to establish the context independence of the obtained effects.

I believe that the debate environment has more to offer than this study has managed to demonstrate. For example, as we saw in Chapter 5, students reported that debating activities and environment hold potential for improving the reading and listening skills. This perception is understandable and quite defendable. Since debate pedagogy positively affects many areas of speaking and writing and largely benefits vocabulary, there is some reason to assume that listening and reading skills experience improvement as well. This is a promising line of research that needs to be explored.

Another relevant research avenue is exploring the interplay between language and argumentation skills. It would be interesting and profitable to research the extent to which the improvement of argumentation skills is affected by the improvement of language skills and vice versa. This is a relevant research avenue that could yield interesting pedagogical insights.

10.6 Concluding remarks

It is important to emphasize that the effectiveness of debate pedagogy hinges on effective implementation. The success of any teaching tool is conditioned by the willingness of learners to launch into its activities with enthusiasm and give it their full effort. Yet, it needs to be noted that successful involvement of learners in the learning processes of debate is largely facilitated by the decision taken by instructors during debate implementation.

This PhD adventure has helped me to broaden my knowledge and understanding of a wide range of important pedagogical topics, a tremendous asset to an L2 teacher like me. The subject of my PhD thesis involved many strands of L2 acquisition and teaching, each of which has offered a mine of invaluable insights into language learning and teaching. I have become more conscious about every move and choice in terms of what I teach (content) and how I teach (pedagogy). In short, I am very satisfied with this research journey that has not only transformed me into a researcher but has also broadened my horizons about teaching a second language as well as providing with me with invaluable insights into its underlying processes.

Indeed, the doctoral grant for teachers in the Netherlands has opened up new avenues in education research (particularly in secondary school contexts) with promising implications for teaching practice. While secondary school instructors may be knowledgeable about the teaching realities of their teaching context, researchers possess the required expertise to navigate the research world. Bridging these two worlds is likely to bring many benefits to pedagogy, and more importantly, bring to the fore new insights into the mechanisms of the acquisition of knowledge and skills. This initiative, therefore, needs continuation because the benefits and rewards are great enough to justify the effort and means invested.

References

Aclan, E. M., & Aziz, N. H. A. (2015a). 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*(2), 1–16. http://doi.org/10.7575/aiac.ijalel.v.4n.2p.1

Aclan, E. M., & Aziz, N. H. A. (2015b). Why and how EFL students learn vocabulary in parliamentary debate class. *Advances in Language and Literary Studies*, *6*(1), 102–113. http://doi.org/10.7575 /aiac.alls.v.6n.1p.102

Adams, R., & Ross-Feldman, L. (2008). Does writing influence learner attention to form? In D. Belcher & A. Hirvela (Eds.), *The oral-literate connection. Perspectives on L2 speaking, writing, and other media interactions* (pp. 243–266). University of Michigan Press. https://doi.org/10.3998/mpub.231182

Ahmadian, M. J. (2012). Task repetition in ELT. Retrieved May 9, 2015, from http://eltj.oxfordjournals.org/-content/early/2012/04/24/elt.ccs020.full

Akerman, R., & Neale, I. (2011). *Debating the evidence: An international review of current situation and perceptions*. CFBT Education Trust.

Alasmari, A., & Ahmed, S. (2013). Using debate in EFL classes English. *Language Teaching*, *6*(1), 147–152. http://doi.org/10.5539/elt.v6n1p147

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*(1), 49–60.

Aldosari, H. S. (2014). The entwined effects of attitude, motivation and gender on EFL learning: A correlation study. *Studies in Literature and Language*, *8*(1), 1–5. http://dx.doi.org/10.3968/n

Alén, E., Domínguez, T., & de Carlos, P. (2015). University students' perceptions of the use of academic debates as a teaching methodology. *Journal of Hospitality, Leisure, Sport & Tourism Education*, *16*, 15–21. https://doi.org/10.1016/j.jhlste.2014.11.001

Alford, K. L., & Surdu, J. R. (2002). Using in-class debates as a teaching tool. *ASEE/IEEE Frontiers in Education Conference*, S1F-10–S1F-15. https://doi.org/10.1109/FIE.2002.1158634

Allen, H. W. (2018). Redefining writing in the foreign language curriculum: Toward a design approach. *Foreign Language Annals*, *51*(3), 513–532. https://doi.org/10.1111/flan.12350

Al-Mahrooqi, R., & Tabakow, M. (2015). Effectiveness of debate in ESL/EFL context courses in the Arabian Gulf: A comparison of two recent student-centered studies in Oman and in Dubai, UAE. In *Proceedings of the Second 21st Century Academic Forum Conference on Teaching, Learning, and Research in the "Just Google It" Age* (pp. 417–428). Harvard University, USA.

Altinyelken, H. K. (2015). *Evolution of curriculum systems to improve learning outcomes and reduce disparities in school achievement* (EFA GMR background report No. ED/EFA/MRT/2015/PI /13). UNESCO. http://unesdoc.unesco.org/images/0023/002324/232420e.pdf

Anderson, J. (2016). Teaching debate in Japanese universities. *The Kwassui Review*, *59*, 73–88.

Anderson, S., & Mezuk, B. (2012). Participating in a policy debate program and academic achievement among at-risk adolescents in an urban public school district: 1997–2007. *Journal of Adolescence*, *35*(5), 1125–1236. https://doi.org/10.1016/j.adolescence.2012.04.005

Archer, C. C., & Miller, M. K. (2011). Prioritizing active learning: An exploration of gateway courses in political science. *PS: Political Science & Politics*, *44*(2), 429–434. https://doi.org/10.1017/S104909651100029

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

Badger, R., & White, G. (2000). A process genre approach to teaching writing. *ELT Journal*, *54*(2), 153–160. https://doi.org/10.1093/elt/54.2.153

Baker, J., & Westrup, H. (2003). *Essential speaking skills: A handbook for English language teachers*. Continuum.

Baker-Smemoe, W., Dewey, D. P., Bown, J., & Martinsen, R. A. (2014). Does measuring L2 utterance fluency equal measuring overall L2 proficiency? Evidence from five languages. *Foreign Language Annals*, *47*(4), 707–728. https://doi.org/10.1111/flan.12110

Bannan-Ritland, B. (2003). The Role of design in research: The integrative learning design. *Framework Educational Researcher*, *32*(1), 21–24. https://doi.org/10.3102/0013189X032001021

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *Journal of the Learning Sciences*, *13*(1), 1–14. https://doi.org/10.1207/s15327809jls1301_1

Baron, J., & Brown, R. V. (Eds.). (1991). *Teaching decision making to adolescents*. Lawrence Erlbaum.

Beeker, A. (2012). Is havo Engels goed genoeg voor het hbo? Een verkennend onderzoek naar mogelijke aansluitproblemen [Is secondary education English level sufficient for higher education? An exploratory study into possible connection problems]. SLO.

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 [ERK-level of writing skills of English, German and French; Research on the level reached at the end of havo and vwo]. *Levende Talen Tijdschrift*, *16*(2), 16–25.

Belcher, D., & Hirvela, A. (Eds.). (2008). *The oral-literate connection: Perspectives on L2 speaking, writing, and other media interactions*. The University of Michigan Press. https://doi.org/10.3998/mpub.231182

Bell, P. (2004). Promoting students' argument construction and collaborative debate in the science classroom. In M. Linn, E. Davis, & P. Bell (Eds.), *Internet environments for science education* (pp. 115–143). Lawrence Erlbaum. https://doi.org/10.4324/9781410610393

Bellon, J. (2000). A research-based justification for debate across the curriculum. *Argumentation and Advocacy*, *36*(3), 161–175.

Benati, A. (2017). The role of input and output tasks in grammar instruction: Theoretical, empirical and pedagogical considerations. *Studies in Second Language Learning and Teaching*, *7*(3), 377–396. https://doi.org/10.14746/ssllt.2017.7.3.2

Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education, 94*(5), 765–793. https://doi.org/10.1002/sce.20402

Bibby, N. (2014). *Discovering the world through debate. A practical guide to educational debate for debaters, coaches and judges.* International Debate Education Association.

Blake, C. (2009). Potential of text-based Internet chats for improving oral fluency in a second language. *The Modern Language Journal, 93*(2), 227–240. https://doi.org/10.1111/j.1540-4781.2009.00858.x

Block, D. (2000). Revisiting the gap between SLA researchers and language teachers. *Links & Letters, 7*, 129–43.

Blumberg, P. (2009). *Developing learner-centred teaching: A practical guide for faculty.* Jossey-Bass. https://doi.org/10.1002/j.0022-0337.2009.73.9.tb04801.x

Boersma, P., & Weenink, D. (2016). *PRAAT: Doing phonetics by computer* [Computer software]. http://www.praat.org/

Bonset, H., & Braaksma, M. (2008). *Het schoolvak Nederlands opnieuw onderzocht: Een inventarisatie van onderzoek van 1997 tot en met 2007* [The school subject of Dutch researched again: An inventory research from 1997 to 2007]. SLO.

Bonwell, C. C., & Eison, J. A. (1991). *Active learning: Creating excitement in the classroom.* Education Resource Information Center.

Bourdillon, M. (2004). Conditions for constructive debate. *South African Journal of Science, 100*, 263–265.

Branham, R. J. (1991). *Debate and critical analysis: The harmony of conflict.* Routledge.

Bransford, J. D., & Schwartz, D. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education, 24*(1), 61–100. https://doi.org/10.3102/0091732X024001061

Breger, B. (2000). Overview of the urban debate program. *Rostrum, 75*(14), 1–2.

Brinton, D. M., Snow, M. A., & Wesche, M. B. (1989). *Content-based second language instruction.* Newbury House.

Brown, H. D. (2004). *Language assessment: Principles and classroom practices.* Pearson Education.

Brown, N. A. (2009). Argumentation and debate in foreign language instruction: A case for the traditional classroom facilitating advanced-level language uptake. *The Modern Language Journal, 93*(4), 534–549. https://doi.org/10.1111/j.1540-4781.2009.00929.x

Brown, T., & Bown, J. (2014). *Teaching advanced language skills through global debate: Theory and practice.* Georgetown University Press.

Brown, Z. (2015). The use of in-class debates as a teaching strategy in increasing students' critical thinking and collaborative learning skills in higher education. *Educational Futures, 7*(1), 39–55. http://hdl.handle.net/2436/621883

Brown, Z., & Wilson, M. (2016). The complexity of in-class debates in higher education: Student perspectives on differing designs. *Educational Futures, 7*(2), 14–28.

Budesheim, T. L., & Lundquist, A. R. (1999). Consider the opposite: Opening minds through in-class debates on course-related controversies. *Teaching of Psychology, 26*(2), 106–110. https://doi.org/10.1207/s15328023top2602_5

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing, 26*, 42–65. https://doi.org/10.1016/j.jslw.2014. 09.005

Butt, N. (2010). *Argument construction, argument evaluation, and decision-making: A content analysis of argumentation and debate textbooks* [Doctoral dissertation, Wayne State University]. https://digitalcommons.wayne.edu/oa_dissertations/77

Bygate, M. (1996). Effects of task repetition: Appraising the developing language of learners. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 136–146). Heinemann.

Bygate, M. (1998). Theoretical perspectives on speaking. *Annual Review of Applied Linguistics, 18*, 20–42. https://doi.org/10.1017/S0267190500003469

Byrd, J. (2008). *Guidebook for student-centered classroom discussions* (S. G. Lea, Ed.). Interactivity Foundation.

Caine, R. N., & Caine, G. (1991). *Making connections: Teaching and the human brain*. Association of Supervision and Curriculum Development.

Camp, J. M., & Schnader, A. L. (2010). Using debate to enhance critical thinking in the accounting classroom: The Sarbanes-Oxley Act and US tax policy. *Issues in Accounting Education, 25*(4), 655–675. https://doi.org/10.2308/iace.2010.25.4.655

Carlin, D., & Payne, J. (1995). *Public speaking today* (2nd ed.). National Textbook Company.

Chapple, L., & Curtis, A. (2000). Content-based instruction in Hong Kong: Student responses to film. *System, 28*(3), 419–433. https://doi.org/10.1016/S0346-251X(00)00021-X

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. https://doi.org/10.1007/978-3-030-01255-7

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. https://doi.org/10.1080/09588221.2011.606224

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. https://doi.org/10.1007/s11191-016-9811-0

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*(2), 100–147. https://doi.org/10.1080/07370008.2016.1145120

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

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

Cho, H. (2017). Synchronous web-based collaborative writing: Factors mediating interaction among second-language writers. *Journal of Second Language Writing, 36*, 37–51. https://doi.org/10.1016/j.jslw.2017.05.013

Cho, Y., & Choi, I. (2018). Writing from sources: Does audience matter? *Assessing Writing, 37*, 25–38. https://doi.org/10.1016/j.asw.2018.03.004

Cinganotto, L. (2019). Debate as a teaching strategy for language learning. *Lingue e Linguaggi, 30*, 107–125. https://doi.org/10.1285/i22390359v30p107

Claxton, N. (2008). *Using deliberative techniques in the English as a foreign language classroom: A manual for teachers of advanced level students*. International Debate Education Association.

Cohen, J. (1988). *Statistical power analyses for the behavioral sciences* (2nd ed.). Lawrence Erlbaum. https://doi.org/10.4324/9780203771587

Cohen, M. (1991). Making class participation a reality. *PS: Political Science & Politics, 24*(4), 699–703. https://doi.org/10.2307/419408

Colbert, K. (1987). The effects of CEDA and NDT debate training on critical thinking ability. *Journal of the American Forensic Association, 23*(24), 194–201. https://doi.org/10.1080/00028533.1987.11951345

Collier, L. (2004, June 16–19). *Argument for success: A study of academic debate in the urban high schools of Chicago, Kansas City, New York, St. Louis and Seattle* [Paper presentation]. International Conference on Social Sciences, Honolulu, Hawaii, United States.

Connor, U. (1990). Linguistic/rhetorical measures for international persuasive student writing. *Research in the Teaching of English, 24*(1), 67–87. https://www.jstor.org/stable/40171446

Council of Europe. (2018). *The common European framework of reference for languages: Learning, teaching, assessment—Companion volume with new descriptors*. Council of Europe Publishing. https://rm.coe.int/cefr-companion-volumewith-new-descriptors-2018/1680787989

Crammond, J. (1998). The uses of complexity of argument structures in expert and student persuasive writing. *Written Communication, 15*(2), 230–268. https://doi.org/10.1177/0741088398015002004

Crookes, G., & Chandler, P. M. (2001). Introducing action research into the education of postsecondary foreign language teachers. *Foreign Language Annals, 34*(2), 131–40. https://doi.org/10.1111/j.1944-9720.2001.tb02818.x

Crossley, S. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research 11*(3), 415–443. https://doi.org/10.17239/jowr-2020.11.03.01

Crossley, S., McNamara, D. S., Baker, R., Wang, Y., Paquette, L., Barnes, T., & Bergner, Y. (2015). Language to completion: Success in an educational data mining massive open online class. In *Proceedings of the 8th International Conference on Educational Data Mining* (pp. 388–392). Madrid, Spain.

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2009). Measuring L2 lexical growth using hypernymic relationships. *Language Learning, 59*(2), 307–334. https://doi.org/10.1111/j.1467-9922.2009.00508.x

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2015). Assessing lexical proficiency using analytic ratings. *Applied Linguistics, 36*(5), 570–590. https://doi.org/10.1093/applin/amt056

Crosson, A. C., & Lesaux, N. K. (2013). Connectives: Fitting another piece of the vocabulary instruction puzzle. *The Reading Teacher, 67*(3), 193–200. https://doi.org/10.1002/TRTR.1197

Crowell, A., & Kuhn, D. (2014). Developing dialogic argumentation skills: A 3-year intervention study. *Journal of Cognition and Development, 15*(2), 363–381. https://doi.org/10.1080/15248372.2012.725187

Cumming, A. (1990). Metalinguistic and ideational thinking in second language composing. *Written Communication, 7*(4), 482–511. https://doi.org/10.1177/0741088390007004003

Darby, M. (2007). Debate: A teaching-learning strategy for developing competence in communication and critical thinking. *Journal of Dental Hygiene, 81*(4), 1–10.

Davis, K. A., Zorwick, M. L. W., Roland, J., & Wade, M. M. (Eds.). (2016). *Using debate in the classroom: Encouraging critical thinking, communication, and collaboration.* Routledge. https://doi.org/10.4324/9781315707808

De Haan, P., & Van Esch, K. (2005). The development of writing in English and Spanish as foreign languages. *Assessing Writing, 10*(2), 100–116. https://doi.org/10.1016/j.asw.2005.05.003

De Haan, P., & Van Esch, K. (2008). Measuring and assessing the development of foreign language writing competence. *Porta Linguarum, 9*, 7–21.

De Jong, N. H. (2016). Fluency in second language assessment. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 203–218). De Gruyter Mouton. https://doi.org/10.1515/9781614513827

De Jong, N. H. (2018). Fluency in second language testing: Insights from different disciplines. *Language Assessment Quarterly, 15*(3), 237–254. https://doi.org/10.1080/15434303.2018.1477780

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 A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 121–142). John Benjamins. https://doi.org/10.1075/lllt.32

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

DeKeyser, R. M. (2007b). Introduction: Situating the concept of practice. In R. M. Dekyser (Ed.), *Practice in a Second Language: Perspectives from applied linguistics and cognitive psychology* (pp. 1–18). Cambridge University Press. https://doi.org/10.1017/CBO9780511667275

Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning, 63*(2), 163–185. https://doi.org/10.1111/lang.12000

Derwing, T. M., Munro, M. J., Foote, J. A., Waugh, E., & Fleming, J. (2014). Opening the window on comprehensible pronunciation after 19 years: A workplace training study. *Language Learning*, *64*(3), 526–548. https://doi.org/10.1111/lang.12053

De Vrind, E. (2020). The SpeakTeach method: Towards self-regulated learning of speaking skills in foreign languages in secondary schools: An adaptive and practical approach [Doctoral dissertation, Leiden University]. ICLON PhD Dissertation Series. https://hdl.handle.net/1887 /138398

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*(6), 679–697. https://doi.org/10.1177/1362168817692161

Dickson, R. (2004). Developing "real-world intelligence": Teaching argumentative writing through debate. *The English Journal*, *94*(1), 34–40. https://doi.org/10.2307/4128845

Do, S. L., & Schallert, D. L. (2004). Emotions and classroom talk: Toward a model of the role of affect in students' experiences of classroom discussions. *Journal of Educational Psychology*, *96*(4), 619–634. https://doi.org/10.1037/0022-0663.96.4.619

Dobbs, C. L. (2014). Signaling organization and stance: Academic language use in middle grade persuasive writing. *Reading and Writing*, *27*, 1327–1352. https://doi.org/10.1007/s11145-013-9489-5

Doody, O., & Condon, M. (2012). Increasing student involvement and learning through using debate as an assessment. *Nurse Education in Practice*, *12*(4), 232–237. https://doi.org/10.1016/j.nepr.2012. 03.002

Dörnyei, Z. (2002). The motivational basis of language learning tasks. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 137–158). John Benjamins. https://doi.org/10.1075/lllt.2

Dörnyei, Z., & Kormos, J. (2000). The role of individual and social variables in oral task performance. *Language Teaching Research*, *4*(3), 275–300. https://doi.org/10.1177/136216880000400305

Dörnyei, Z., & Ryan, S. (2015). *The psychology of the second language learner revisited.* Routledge. https://doi.org/10.4324/9781315779553

Dörnyei, Z., & Taguchi, T. (2009). *Questionnaires in second language research: Construction, administration, and processing.* Routledge. https://doi.org/10.4324/9780203864739

Driscoll, W., & Zompetti, J. P. (2003). *Discovering the world through debate: A practical guide to educational debate for debaters, coaches, and judges.* International Debate Education Association.

Dundes, L. (2001). Small group debates: Fostering critical thinking in oral presentations with maximal class involvement. *Teaching Sociology*, *29*(2), 237–243. https://doi.org/10.2307/1318721

Duran-Karaoz, Z., & Tavakoli, P. (2020). Predicting L2 fluency from L1 fluency behavior: The case of L1 Turkish and L2 English speakers. *Studies in Second Language Acquisition*, *42*(4), 671–695. https://doi.org/10.1017/S0272263119000755

Dy-Boarman, E. A., Bryant, G. A., Herring, M. S., & Foster, K. Y. (2018). Impact of debates on student perceptions and competency scores in the advanced pharmacy practice setting. *Currents in Pharmacy Teaching and Learning*, *10*(1), 66–71. https://doi.org/10.1016/j.cptl.2017.09.011

East, M. (2012). *Task-based language teaching from the teachers' perspective*. John Benjamins. https://doi.org/10.1075/tblt.3

Ellis, R. (1997a). SLA and language pedagogy: An educational perspective. *Studies in Second Language Acquisition*, *19*(1), 69–92. https://doi.org/10.1017/S0272263197001058

Ellis, R. (1997b). *SLA research and language teaching*. Oxford University Press.

Ellis, R. (1997c). The empirical evaluation of language teaching materials. *ELT journal*, *51*(1), 36–42. https://doi.org/10.1093/elt/51.1.36

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Ellis, R. (2005). *Instructed second language acquisition: A literature review*. New Zealand Ministry of Education.

Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics*, *30*(4), 474–509. https://doi.org/10.1093/applin/amp042

Ellis, R. (2010). Second language acquisition, teacher education and language pedagogy. *Language Teaching*, *43*(2), 182–201. https://doi.org/10.1017/S0261444809990139

Ellis, R. (2012). *Language teaching research and language pedagogy*. Wiley Blackwell. https://doi.org/10.1002/9781118271643

Ellis, R. (2018). Taking the critics to task: The case for task-based teaching. In I. Walker, D. Chan, M. Nagami, & C. Bourguignon (Eds.), *New perspectives on the development of communicative and related competence in foreign language education* (pp. 23–40). De Gruyter Mouton. https://doi.org/10.1515/9781501505034-002

Ellis, R. (2021). A short history of SLA: Where have we come from and where are we going? *Language Teaching*, *54*(2), 190–205. https://doi.org/10.1017/S0261444820000038

Ellis, R., & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition research*. Routledge. https://doi.org/10.4324/9780203796580

Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, *26*(1), 59–84. https://doi.org/10.1017/S0272263104026130

El Majidi, A., de Graaff, R., & Janssen, D. (2015a). Debatteren in het Engels. Leerzame ingrediënten in de ogen van leerlingen [Debating in English. Instructive ingredients in the eyes of students]. *Levende Talen Tijdschrift*, *16*(1), 24–33.

El Majidi, A., de Graaff, R., & Janssen, D. (2015b). Invest in what energizes students to learn: Investigating students' attitude towards debate in the foreign language classroom. *Journal of Language Teaching and Research*, *6*(5), 924–932. https://doi.org/10.17507/jltr.0605.03

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*(1), 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*(4), 804–821. https://doi.org/10.1111/modl.12673

El Majidi, A., de Graaff, R., & Janssen, D. (in press). Debate as a pedagogical tool for developing speaking skills in second language education. *Language Teaching Research*. https://doi.org/10.1177/13621688211050619

El Majidi, A., Janssen, D., & de Graaff, R. (2021). The effects of in-class debates on argumentation skills in second language education. *System*, *101*, 1–15. https://doi.org/10.1016/j.system.2021.102576

Emaliana, I. (2017). Teacher-centered or student-centered learning approach to promote learning? *Jurnal Sosial Humaniora*, *10*(2), 59–70. http://dx.doi.org/10.12962/j24433527.v10i2.2161

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education*, *88*(6), 915–933. https://doi.org/10.1002/sce.20012

Ericson, J. M., Murphy, J. J., & Zeuschner, R. B. (2003). *The debater's guide* (3rd ed.). South Illinois University Press.

Fallahi, C. R., & Haney, C. R. (2007). Using debate in helping students discuss controversial topics. *Journal of College Teaching & Learning*, *4*(10), 83–88. https://doi.org/10.19030/tlc.v4i10.1540

Fasoglio, D., & Tuin, D. (2017). *Eindniveau gespreksvaardigheid Engels in het voortgezet onderwijs* [Attainment level of English speaking conversational skills in secondary education]. SLO.

Faust, J. L., & Paulson, D. R. (1998). Active learning in the college classroom. *Journal on Excellence in College Teaching*, *9*(2), 3–24.

Ferretti, R. P., & Graham, S. (2019). Argumentative writing: Theory, assessment, and instruction. *Reading and Writing*, *32*(6), 1345–1357. https://doi.org/10.1007/s11145-019-09950-x

Ferretti, R. P., MacArthur, C. A., & Dowdy, N. S. (2000). The effects of an elaborated goal on the persuasive writing of students with learning disabilities and their normal achieving peers. *Journal of Educational Psychology*, *92*(4), 694–702. https://doi.org/10.1037/0022-0663.92.4.694

Ferris, D. R. (2011). *Treatment of error in second language student writing* (2nd ed.). University of Michigan Press. https://doi.org/10.3998/mpub.2173290

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. https://doi.org/10.1016/S1060-3743(01)00039-X

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*(3), 354–375. https://doi.org/10.1093/applin/21.3.354

Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition*, *26*(2), 275–301. https://doi.org/10.1017/S0272263104262064

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

Gadzella, B. M., Hartsoe, K., & Harper, J. (1989). Critical thinking and mental ability groups. *Psychological Reports*, *65*(3), 1019–1026. https://doi.org/10.2466/pr0.1989.65.3.1019

Gardner, R. C. (1985). *Social psychology and language learning: The role of attitudes and motivation.* Edward Arnold.

Garrett, M., Schoener, L., & Hood, L. (1996). Debate: A teaching strategy to improve verbal communication and critical-thinking skills. *Nurse Education, 21*(4), 37–40.

Gass, S., & Mackey, A. (2015). Input, interaction and output in second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 180–206). Routledge. https://doi.org/10.4324/9780203628942

Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to fluency. *Canadian Modern Language Review, 61*(3), 325–353. https://doi.org/10.3138 /cmlr.61.3.325

Gelinck, C. (2000). Discussie gesloten? Het debat op school—Een betoog [Debate closed? The debate at school—An essay]. *Levende Talen Tijdschrift, 1*(1), 21–26.

Goh, C. C. (2017). Research into practice: Scaffolding learning processes to improve speaking performance. *Language Teaching, 50*(2), 247–260. https://doi.org/10.1017/S0261444816000483

Goodwin, J. (2003). Students' perspectives on debate exercises in content area classes. *Communication Education, 52*(2), 157–163. https://doi.org/10.1080/03634520302466

Gravemeijer, K., & Cobb, P. (2006). Design research from a learners' perspective. In J. van den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), *Educational design research* (pp. 17–51). Routledge.

Green, J. H. (2015). Teaching for transfer in EAP: Hugging and bridging revisited. *English for Specific Purposes, 37*, 1–12. https://doi.org/10.1016/j.esp.2014.06.003

Griffee, D. T. (2012). *An introduction to second language research methods: Design and data.* TESL-EJ Publications.

Gulnaz, F. (2020). Fostering Saudi EFL learners' communicative, collaborative and critical thinking skills through the technique of in-class debate. *International Journal of English Linguistics, 10*(5), 265–283. https://doi.org/10.5539/ijel.v10n5p265

Haijma, A. (2013). Duiken in een taalbad; onderzoek naar het gebruik van doeltaal als voertaal [Immersed in a language bath; research into the use of the target language as the language of instruction]. *Levende Talen Tijdschrift, 14*(3), 27–40.

Hall, D. (2011). Debate: Innovative teaching to enhance critical thinking and communication skills in healthcare professionals. *The Internet Journal of Allied Health Sciences and Practice, 9*(3), 1–8. https://doi.org/10.46743/1540-580X/2011.1361

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

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

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

Hill, B. (1982). Intercollegiate debate: Why do students bother? Southern speech. *Communication Journal, 48*(1), 77–88. https://doi.org/10.1080/10417948209372554

Hirvela, A. (2013). Preparing English language learners for argumentative writing. In L. C. de Oliveira & T. J. Silva (Eds.), *L2 writing in secondary classrooms: Student experiences, academic issues, and teacher education* (pp. 67–86). Routledge. https://doi.org/10.4324/9780203082669

Hirvela, A. (2016). *Connecting reading and writing in second language writing instruction* (2nd ed.). The University of Michigan Press. https://doi.org/10.3998/mpub.8122864

Hirvela, A. (2017). Argumentation & second language writing: Are we missing the boat? *Journal of Second Language Writing, 36*, 69–74. https://doi.org/10.1016/j.jslw.2017.05.002

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. https://doi.org/10.1016/j.jeap.2018.02.001

Hoffmann, M. H. (2016). Reflective argumentation: A cognitive function of arguing. *Argumentation, 30*(4), 365–397. https://doi.org/10.1007/s10503-015-9388-9

Hsu, H. C. (2019). The combined effect of task repetition and post-task transcribing on L2 speaking complexity, accuracy, and fluency. *The Language Learning Journal, 47*(2), 172–187. https://doi.org/10.1080/09571736.2016.1255773

Hsu, P. S., Van Dyke, M., Chen, Y., & Smith, T. J. (2015). The effect of a graph-oriented computer-assisted project-based learning environment on argumentation skills. *Journal of Computer Assisted Learning, 31(1)*, 32–58. https://doi.org/10.1111/jcal.12080

Hsu, P. S., Van Dyke, M., & Smith, T. J. (2017). The effect of varied gender groupings on argumentation skills among Middle school students in different cultures. *Middle Grades Review, 3*(2), 1–22. https://scholarworks.uvm.edu/mgreview/vol3/iss2/4

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 & Writing Quarterly, 36*(4), 339–364. https://doi.org/10.1080/10573569.2019.1649223

Huensch, A., & Tracy–Ventura, N. (2017). L2 utterance fluency development before, during, and after residence abroad: A multidimensional investigation. *The Modern Language Journal, 101*(2), 275–293. https://doi.org/10.1111/modl.12395

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

Hwang, H., Jung, H., & Kim, H. (2020). Effects of written versus spoken production modalities on syntactic complexity measures in beginning-level child EFL learners. *The Modern Language Journal, 104*(1), 267–283. https://doi.org/10.1111/modl.12626

Hyde, C. A., & Ruth, B. J. (2002). Multicultural content and class participation: Do students self-censor? *Journal of Social Work Education, 38*(2), 241–256. https://doi.org/10.1080/10437797.2002.10779095

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. https://doi.org/10.3998/mpub.231182

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

Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing, 16*, 148–164. https://doi.org/10.1016/j.jslw.2007.07.005

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). John Benjamins. https://doi.org/10.1075/lllt.31

Hyland, K. (2019). *Second language writing* (2nd ed.). Cambridge University Press. https://ueaeprints.uea.ac.uk/id/eprint/69309

Iberri-Shea, G. (2013). Teaching English through debate in classroom contexts. In A. L. Sellami (Ed.), *Proceedings of the 2013 4th International Conference on Argumentation, Rhetoric, Debate, and Pedagogy,* (pp. 129–136). Qatar Debate Center.

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*(4), 487–505. https://doi.org/10.1080/09571736.2015.1130079

Inoue, N., & Nakano, M. (2004, June 27–30). *The benefits and costs of participating in competitive debate activities: Differences between Japanese and American college students* [Paper presentation]. Venice Argumentation Conference, Wake Forest University, USA.

Iordanou, K. (2013). Developing face-to-face argumentation skills: Does arguing on the computer help? *Journal of Cognition and Development, 14*(2), 292–320. https://doi.org/10.1080/15248372.2012.668732

Ivey, G. (2013). Developing an intervention to increase engaged reading among adolescents. In T. Plomp & N. Nieveen (Eds.), *Educational design research—Part B: Illustrative cases* (pp. 235–251). SLO.

Izumi, S., Bigelow, M., Fujiwara, M., & Fearnow, S. (1999). Testing the output hypothesis: Effects of output on noticing and second language acquisition. *Studies in Second Language Acquisition, 21*(3), 421–452. https://doi.org/10.1017/S0272263199003034

Jackson, M. (1973). Debate: A neglected teaching tool. *Peabody Journal of Education, 50*(2), 150–154.

Jansen, M., Lüdtke, O., & Schroeders, U. (2016). Evidence for a positive relation between interest and achievement: Examining between-person and within-person variation in five domains. *Contemporary Educational Psychology, 46*, 116–127. https://doi.org/10.1016/j.cedpsych.2016.05.004

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. https://doi.org/10.1016/j.jslw.2018.06.005

Jiménez, J. L., Perdiguero, J., & Suárez, A. (2011). Debating as a classroom tool for adapting learning outcomes to the European higher education area. Retrieved July 29, 2014, from http://www.ub.edu/irea/working_papers/2011/201109.pdf

Jin, T., Su, Y., & Lei, J. (2020). Exploring the blended learning design for argumentative writing. *Language Learning & Technology, 24*(2), 23–34. http://hdl.handle.net/10125/44720

Johnson, B. (2003). *The student-centered classroom handbook: A guide to implementation.* Eye on Education, Inc.

Johnson, R. H. (2013). The role of audience in argumentation from the perspective of informal logic. *Philosophy & Rhetoric, 46*(4), 533–549. https://doi.org/10.5325/philrhet.46.4.0533

Jonassen, D. H., & Kim, B. (2010). Arguing to learn and learning to argue: Design justifications and guidelines. *Education Technology Research Development, 58*, 439–457. https://doi.org/10.1007/s11423-009-9143-8

Jones, K. T. (1994). Cerebral gymnastics 101: Why do debaters debate? *CEDA Yearbook, 15*, 65–75.

Jönsson, S., & Lukka, K. (2007). There and back again: Doing interventionist research in management accounting. In C. S. Chapman, A. G. Hopwood, & M. D. Shields (Eds.), *Handbook of management accounting research* (pp. 373–398). Elsevier.

Kang, S. (2005). Dynamic emergence of situational willingness to communicate in a second language. *System, 33*(2), 277–292. https://doi.org/10.1016/j.system.2004.10.004

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

Kathpalia, S. S., & See, E. K. (2016). Improving argumentation through student blogs. *System, 58*, 25–36. https://doi.org/10.1016/j.system.2016.03.002

Kayi, H. (2006). Teaching speaking: Activities to promote speaking in a second language. Retrieved February 2, 2015, from http://iteslj.org/Techniques/Kayi-TeachingSpeaking.html

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

Kennedy, R. (2009). The power of in-class debates. *Active learning in higher education, 10*(3), 225–236. https://doi.org/10.1177/1469787409343186

Khan, S. A., Omar, H., Babar, M. H., & Toh, C. G. (2012). Utilization of debate as an educational tool to learn health economics for dental students in Malaysia. *Journal of Dental Education, 76*(12), 1675–1683. https://doi.org/10.1002/j.0022-0337.2012.76.12.tb05431.x

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. https://doi.org/10.20802/eibeibunka.28.0_23

Klein, W. (1998). The contribution of second language acquisition research. *Language Learning, 48*(4), 527–49. https://doi.org/10.1111/0023-8333.00057

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. https://doi.org/10.1016/j.asw.2014.01.001

Kopcha, T. J., McGregor, J., Shin, S., Qian, Y., Choi, J., Hill, R., Mativo, J., & Choi, I. (2017). Developing an integrative STEM curriculum for robotics education through educational design research. *Journal of Formative Design in Learning, 1*(1), 31–44. https://doi.org/10.1007/s41686-017-0005-1

Kormos, J. (2014). Differences across modalities of performance: An investigation of linguistic and discourse complexity in narrative tasks. In H. Byrnes & R. M. Manchón (Eds.), *Task-based language learning—Insights from and for L2 writing* (pp. 193–217). John Benjamins. https://doi.org/10.1075/tblt.7

Kormos, J., & Préfontaine, Y. (2017). Affective factors influencing fluent performance: French learners' appraisals of second language speech tasks. *Language Teaching Research, 21*(6), 699–716. https://doi.org/10.1177/1362168816683562

Kouwenhoven, W. (2013). Towards a competence-based curriculum for a new faculty of education of the Eduardo Mondlane University, Mozambique: A reconstructive study. In T. Plomp & N. Nieveen (Eds.), *Educational design research—Part B: Illustrative cases* (pp. 1037–1059). SLO.

Krahnke, K. (1987). *Approaches to syllabus design for foreign language teaching.* Prentice Hall.

Krieger, D. (2005). Teaching debate to ESL students: A six-class unit. *The Internet TESL Journal, 11(2)*, 25–45.

Kuhn, D. (1991). *The skills of argument.* Cambridge University Press. https://doi.org/10.1017/CBO9780511571350

Kuhn, D. (2001). Why development does (and doesn't) occur: Evidence from the domain of inductive reasoning. In R. Siegler & J. McClelland (Eds.), *Mechanisms of cognitive development: Neural and behavioral perspectives* (pp. 221–249). Lawrence Erlbaum.

Kuhn, D. (2007). Is direct instruction an answer to the right question? *Educational Psychologist, 42*(2), 109–113. https://doi.org/10.1080/00461520701263376

Kuhn, D. (2018). A role for reasoning in a dialogic approach to critical thinking. *Topoi, 37*(1), 121–128. https://doi.org/10.1007/s11245-016-9373-4

Kuhn, D., & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents' thinking. *Psychological Science, 22*(4), 545–552. https://doi.org/10.1177/0956797611402512

Kuhn, D., Hemberger, L., & Khait, V. (2016). *Argue with me: Argument as a path to developing students' thinking and writing* (2nd ed.). Routledge. https://doi.org/10.4324/9781315692722

Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development, 74*(5), 1245–1260. https://doi.org/10.1111/1467-8624.00605

Kuiken, F., & Vedder, I. (2017). Functional adequacy in L2 writing: Towards a new rating scale. *Language Testing, 34*(3), 321–336. https://doi.org/10.1177/0265532216663991

Kwon, O. N., Ju, M. K., Kim, R. Y., Park, J. H., & Park, J. S. (2013). Design research as an inquiry into students' argumentation and justification: Focusing on the design of intervention. In T. Plomp & N. Nieveen (Eds.), *Educational design research—Part B: Illustrative cases* (pp. 199–220). SLO.

Lambert, C. (2017). Tasks, affect and second language performance. *Language Learning*, *21*(6), 657–664. https://doi.org/10.1177/1362168817736644

Lambert, C., Philp, J., & Nakamura, S. (2017). Learner-generated content and engagement in second language task performance. *Language Teaching Research*, *21*(6), 665–680. https://doi.org/10.1177/1362168816683559

Lee, J. J., & Subtirelu, N. C. (2015). Metadiscourse in the classroom: A comparative analysis of EAP lessons and university lectures. *English for Specific Purposes*, *37*, 52–62. https://doi.org/10.1016/j.esp.2014.06.005

Lee, S., & Pulido, D. (2017). The impact of topic interest, L2 proficiency, and gender on EFL incidental vocabulary acquisition through reading. *Language Teaching Research*, *21*(1), 118–135. https://doi.org/10.1177/1362168816637381

Leki, I., Cumming, A., & Silva, T. (2010). *A synthesis of research on second language writing in English*. Routledge. https://doi.org/10.4324/9780203930250

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press. https://hdl.handle.net/2066/15392

Li, H. H., Zhang, L. J., & Parr, J. M. (2020). Small-group student talk before individual writing in tertiary English writing classrooms in China: Nature and insights. *Frontiers in Psychology*, *11*, 1–14. https://doi.org/10.3389/fpsyg.2020.570565

Li, L., Chen, J., & Sun, L. (2015). The effects of different lengths of pretask planning time on L2 learners' oral test performance. *TESOL Quarterly*, *49*(1), 38–66. https://doi.org/10.1002/tesq.159

Lieb, M. (2007). Teaching debate skills to intermediate and lower level EFL students. Retrieved February 18, 2014, from http://www.tht-japan.org/proceedings/2007/m_lieb73-84.pdf

Lightbown, P. (2000). Anniversary article: Classroom SLA research and second language teaching. *Applied Linguistics*, *21*(4), 431–62. https://doi.org/10.1093/applin/21.4.431

Littlefield, R. S. (2001). High school student perceptions of the efficacy of debate participation. *Journal of the American Forensic Association*, *38*(2), 83–97. https://doi.org/10.1080/00028533.2001.11821559

Liu, F., & Stapleton, P. (2014). Counterargumentation and the cultivation of critical thinking in argumentative writing: Investigating washback from a high-stakes test. *System*, *45*, 117–128. https://doi.org/10.1016/j.system.2014.05.005

Liu, F., & Stapleton, P. (2020). Counterargumentation at the primary level: An intervention study investigating the argumentative writing of second language learners. *System*, *89*, 1–15. https://doi.org/10.1016/j.system.2019.102198

Loewen, S. (2020). *Introduction to instructed second language acquisition* (2nd ed.). Routledge. https://doi.org/10.4324/9781315616797

Loewen, S., & Sato, M. (2018). Interaction and instructed second language acquisition. *Language Teaching*, *51*(3), 285–329. https://doi.org/10.1017/S0261444818000125

Long, M. H. (1985). A role for instruction in second language acquisition: Task-based language training. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and assessing second language acquisition* (pp. 77–99). Multilingual Matters.

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39–52). John Benjamin. https://doi.org/10.1075/sibil.2

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). Academic Press.

Long, M. H. (2009). Methodological principles for language teaching. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 373–394). Wiley-Blackwell. https://doi.org/10.1002/9781444315783

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Wiley Blackwell.

Long, M. H. (2018). Interaction in L2 classrooms. In J. I. Liontas (Ed.), *The TESOL encyclopedia of English language teaching* (pp. 1–7). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118784235.eelt0233

Long, M. H., & Porter, P. A. (1985). Group work, interlanguage talk, and second language acquisition. *TESOL Quarterly*, *19*(2), 207–227. https://doi.org/10.2307/3586827

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, *15*(4), 474–496. https://doi.org/10.1075/ijcl.15.4.02lu

Lubetsky, M., LeBeau, C., & Harrington, D. (2000). *Discover debate: Basic skills for supporting and refuting opinions*. Language Solutions.

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*(1), 18–30.

Lys, F. (2013). The development of advanced learner oral proficiency using iPads. *Language Learning & Technology*, *17*(3), 94–116. http://dx.doi.org/10125/44341

MacArthur, C. A., Ferretti, R. P., & Okolo, C. M. (2002). On defending controversial viewpoints: Debates of sixth graders about the desirability of early 20th-century American immigration. *Learning Disabilities Research & Practice*, *17*(3), 160–172. https://doi.org/10.1111/1540-5826.00042

MacIntyre, P. D., MacMaster, K., & Baker, S. C. (2001). The convergence of multiple models of motivation for second language learning: Gardner, Pintrich, Kuhl, and McCroskey. In Z. Dörnyei & R. Schmidt (Eds.), *Motivation and second language acquisition* (pp. 461–492). Second Language Teaching & Curriculum Center.

Mackey, A. (1999). Input, interaction and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition*, *21*(4), 557–587. https://doi.org/10.1017/S0272263199004027

Malloy, J. A., Tracy, K. N., Scales, R. Q., Menickelli, K., & Scales, W. D. (2020). It's not about being right: Developing argument through debate. *Journal of Literacy Research*, *52*(1), 79–100. https://doi.org/10.1177/1086296X19896495

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). John Benjamins. https://doi.org/10.1075/lllt.31

Manchón, R. M., & Roca de Larios, J. R. (2007). Writing-to-learn in instructed language learning contexts. In E. Alcón & P. Safont (Eds.), *Intercultural language use and language learning* (pp. 101–121). Springer.

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). De Gruyter Mouton. https://doi.org/10.1515/9781614511335

Martens, E. A. (2007). The instructional use of argument across the curriculum. *Middle School Journal*, *38*(5), 4–13. https://doi.org/10.1080/00940771.2007.11461597

Marttunen, M., & Laurinen, L. (2001). Learning of argumentation skills in networked and face-to-face environments. *Instructional Science*, *29*(2), 127–153. https://doi.org/10.1023/A:1003931514884

Masgoret, A. M., & Gardner, R. C. (2003). Attitudes, motivation, and second language learning: A meta-analysis of studies conducted by Gardner and associates. *Language Learning*, *53*(1), 123–163. https://doi.org/10.1111/1467-9922.00227

Matos, F. (2021). Collaborative writing as a bridge from peer discourse to individual argumentative writing. *Reading and Writing, 34*(5), 1321–1342. https://doi.org/10.1007/s11145-020-10117-2

McKenney, S. (2001). *Computer-based support for science education materials developers in Africa: Exploring potentials* [Doctoral dissertation, University of Twente, the Netherlands]. https://research.utwente.nl/en/publications/computer-based-support-for-science-education-materials-developers

McKenney, S., & Reeves, T. C. (2012). *Conducting educational design research*. Routledge. https://doi.org/10.4324/9780203818183

McKenney, S., & Reeves, T. C. (2014). Educational design research. In M. Spector, M. Merrill, J. Elen, & M. Bischop (Eds.), *Handbook of research on educational communications & technology* (pp. 131–140). Springer. https://doi.org/10.1007/978-1-4614-3185-5

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, *27*(1), 57–86. https://doi.org/10.1177/0741088309351547

Means, M. L., & Voss, J. F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. *Cognition and Instruction*, *14*(2), 139–178. https://doi.org/10.1207/s1532690xci1402_1

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, *20*(1), 83–108. https://doi.org/10.1017/S0272263198001041

Mei, W. S. (2006). Creating a contrastive rhetorical stance: Investigating the strategy of problematization in students' argumentation. *RELC Journal, 37*(3), 329–353. https://doi.org/10.1177/0033688206071316

Mercier, H. (2011). Reasoning serves argumentation in children. *Cognitive Development, 26*(3), 177–191. https://doi.org/10.1016/j.cogdev.2010.12.001

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34*(2), 57–74. https://doi.org/10.1017/S0140525X10000968

Merrell, B., Calderwood, K. J., & Flores, R. (2015). The intersection of competitions and classrooms in forensics pedagogy. *Communication Studies, 66*(4), 433–447. https://doi.org/10.1080/10510974.2015.1057751

Merrell, B., Calderwood, K. J., & Graham, T. (2017). Debate across the disciplines: Structured classroom debates in interdisciplinary curricula. *Contemporary Argumentation & Debate, 37*, 57–74.

Meyers, C., & Jones, T. (1993). *Promoting active learning: Strategies for the college classroom.* Jossey-Bass.

Mezuk, B., Bondarenko, I., Smith, S., & Tucker, E. (2011). Impact of participating in a policy debate program on academic achievement: Evidence from the Chicago urban debate league. *Educational Research and Reviews, 6*(9), 622–635. https://doi.org/10.5897/ERR.9000192

Michaels, S., O'Connor, C., & Resnick, L. (2008). Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. *Studies in Philosophy and Education, 27*(4), 283–297. https://doi.org/10.1007/s11217-007-9071-1

Michel, M. (2017). Complexity, accuracy and fluency in L2 production. In S. Loewen & S. Masatoshi (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 50–68). Routledge. https://doi.org/10.4324/9781315676968

Michel, M., Vidon, C., de Graaff, R., & Lowie, W. (2021). Language learning beyond English in the Netherlands: A fragile future? *European Journal of Applied Linguistics, 9*(1), 159–182. https://doi.org/10.1515/eujal-2020-0020

Midgette, E., Haria, P., & MacArthur, C. (2008). The effects of content and audience awareness goals for revision on the persuasive essays of fifth-and eighth-grade students. *Reading and Writing, 21*(1), 131–151. https://doi.org/10.1007/s11145-007-9067-9

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

Mirra, N., Honoroff, B., Elgendy, 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), 1–22.

Mitchell, G. (1998). Pedagogical possibilities for argumentative agency in academic debate. *Argumentation & Advocacy, 35*(2), 41–60. https://doi.org/10.1080/00028533.1998.11951620

Morse, K. (2011). Debate: A tool for language learning. *Journal of Border Educational Research, 10*, 108–119.

Moser, J. (2012). *A classroom-based, mixed methods study into the influence of transcribing, reporting, and task repetition* [Unpublished doctoral dissertation]. University of Birmingham.

Musselman, E. G. (2004). Using structured debate to achieve autonomous student discussion. *The History Teacher, 37*(3), 335–349. https://doi.org/10.2307/1555673

Naghdipour, B. (2016). English writing instruction in Iran: Implications for second language writing curriculum and pedagogy. *Journal of Second Language Writing, 32*, 81–87. https://doi.org/10.1016/j.jslw.2016.05.001

Nakahama, Y., Tyler, A., & Van Lier, L. (2001). Negotiation of meaning in conversational and information gap activities: A comparative discourse analysis. *TESOL Quarterly, 35*(3), 377–405. https://doi.org/10.2307/3588028

Nassaji, H. (2012). The relationship between SLA research and language pedagogy: Teachers' perspectives. *Language Teaching Research 16*(3) 337–365. https://doi.org/10.1177/1362168812436903

Newell, G. E., Beach, R., Smith, J., & VanDerHeide, J. (2011). Teaching and learning argumentative reading and writing: A review of research. *Reading Research Quarterly, 46*(3), 273–304. https://doi.org/10.1598/RRQ.46.3.4

Nieveen, N. (1999). Prototyping to reach product quality. In J. van den Akker, R. M. Branch, K. Gustafson, N. Nieveen, & T. Plomp (Eds.), *Design approaches and tools in education and training* (pp. 125–136). Kluwer Academic.

Nieveen, N. (2007). Formative evaluation in educational design research. In T. Plomp & N. Nieveen (Eds.), *An introduction in educational design research* (pp. 89–102). SLO.

Nisbett, R. (2003). *The geography of thought: How Asians and westerners think differently ... and why.* The Free Press.

Niu, R. (2009). Effect of task-inherent production modes on EFL learners' focus on form. *Language Awareness, 18*(3–4), 384–402. https://doi.org/10.1080/09658410903197256

Noble, W. (2010). Understanding metadiscoursal use: Lessons from a 'local' corpus of learner academic writing. *Nordic Journal of English Studies, 9*(2), 145–169.

Norris, J. M. (2009). Task-based teaching and testing. In M. H. Long & C. Doughty (Eds.), *The handbook of language teaching* (pp. 578–594). Blackwell. https://doi.org/10.1002/9781444315783

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*(4), 555–578. https://doi.org/10.1093/applin/amp044

Nunan, D. (2004). *Task-based language teaching.* Cambridge University Press. https://doi.org/10.1017/CBO9780511667336

Nussbaum, E. M., Kardash, C. M., & Graham, S. E. (2005). The Effects of goal instructions and text on the generation of counterarguments during writing. *Journal of Educational Psychology, 97*(2), 157–169. https://doi.org/10.1037/0022-0663.97.2.157

Nussbaum, E. M., & Schraw, G. (2007). Promoting argument-counterargument integration in students' writing. *The Journal of Experimental Education, 76*(1), 59–92. https://doi.org/10.3200/JEXE.76.1.59-92

O'Keefe, D. J. (1999). How to handle opposing arguments in persuasive messages: A meta-analytic review of the effects of one-sided and two-sided messages. *Annals of the International Communication Association, (22)*1, 209–249. https://doi.org/10.1080/23808985.1999.11678963

O'Mahoney, A. (2015). The ancient art of "debating" and its usefulness in the foreign language classroom: The case of the French debating association. *Journal of Language and Cultural Education, 3*(1), 144–151.

Omelicheva, M. Y. (2006). Global politics on trial: Using educational debate for teaching controversies of world affairs. *International Studies Perspectives, 7*(2), 172–186. https://doi.org/10.1111/j.1528-3585.2006.00238.x

Omelicheva, M. Y., & Avdeyeva, O. (2008). Teaching with lecture or debate? Testing the effectiveness of traditional versus active learning methods of instruction, *PS: Political Science and Politics, 41*(3), 603–607. https://doi.org/10.1017/S1049096508080815

Oros, A. L. (2007). Let's debate: Active learning encourages student participation and critical thinking. *Journal of Political Science Education, 3*(3), 293–311. https://doi.org/10.1080/15512160701558273

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics, 24*(4), 492–518. https://doi.org/10.1093/applin/24.4.492

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). John Benjamins. https://doi.org/10.1075/lllt.31

Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics, 25*, 26–45. https://doi.org/10.1017/S0267190505000024

Osana, H. P., & Seymour, J. R. (2004). Critical thinking in preservice teachers: A rubric for evaluating argumentation and statistical reasoning. *Educational Research and Evaluation, 10*(4–6), 473–498. https://doi.org/10.1080/13803610512331383529

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching, 41*(10), 994–1020. https://doi.org/10.1002/tea.20035

Paek, J. K., & Kang, Y. (2017). Investigation of content features that determine Korean EFL learners' argumentative writing qualities. *English Teaching, 72*(2), 101–122. https://doi.org/10.15858/engtea.72.2.201706.101

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics, 30*(4), 590–601. https://doi.org/10.1093/applin/amp045

Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research, 31*(1), 117–134. https://doi.org/10.1177/0267658314536435

Pally, M. (1997). Critical thinking in ESL: An argument for sustained content. *Journal of Second Language Writing, 6*(3), 293–311. https://doi.org/10.1016/S1060-3743(97)90016-3

Park, C., Kier, C., & Jugdev, K. (2011). Debate as a teaching strategy in online education: A case study. *Canadian Journal of Learning and Technology, 37*(3), 1–17. https://doi.org/10.21432/T2FW2R

Peng, H. (2016). Learner perceptions of Chinese EFL college classroom environments. *English Language Teaching ELT, 9*(1), 22–32. https://doi.org/10.5539/elt.v9n1p22

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. https://doi.org/10.1016/j.jslw.2017.10.013

Phung, L. (2017). Task preference, affective response, and engagement in L2 use in a US university context. *Language Teaching Research, 21*(6), 751–766. https://doi.org/10.1177/1362168816683561

Pica, T. (1996). The essential role of negotiation in the second language classroom. *JALT Journal, 18*(2) 241–68.

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*(2), 301–338. https://doi.org/10.1017/S027226310606013X

Piggott, L. (2019). *First meaning then form: A longitudinal study on the effects of delaying and reducing explicit form-focused instruction for young adolescent EFL learners* (Publication No. 545) [Doctoral dissertation, Utrecht University, the Netherlands]. LOT Dissertation Series. https://dspace.library.uu.nl/handle/1874/386767

Plakans, L., Gebril, A., & Bilki, Z. (2016). Shaping a score: Complexity, accuracy, and fluency in integrated writing performances. *Language Testing, 36*(2), 161–179. https://doi.org/10.1177/0265532216669537

Plomp, T. (2007). Educational design research: An introduction. In T. Plomp & N. Nieveen (Eds.), *An introduction to educational design research* (pp. 9–35). SLO.

Plomp, T., & Nieveen, N. (2013). Introduction to the collection of illustrative cases of educational design research. In T. Plomp & N. Nieveen (Eds.), *Educational design research—Part B: Illustrative cases* (pp. V–XX). SLO.

Polio, C., & Park, J. H. (2016). Language development in second language writing. In R. Manchón & P. K. Matsuda (Eds.), *The handbook of second and foreign language writing* (pp. 287–306). Routledge. https://doi.org/10.1515/9781614511335

Prabhu, N. S. (1987). *Second language pedagogy.* Oxford University Press.

Qin, J., & Karabacak, E. (2010). The analysis of Toulmin elements in Chinese EFL university argumentative writing. *System, 38*(3), 444–456. https://doi.org/10.1016/j.system.2010.06.012

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. https://doi.org/10.1016/j.jslw.2016.06.001

Qiu, X., & Lo, Y. (2017). Content familiarity, task repetition and Chinese EFL learners' engagement in second language use. *Language Teaching Research, 21*(6), 681–698. https://doi.org/10.1177/1362168816684368

Randolph, D. S. (2007). Student perceptions of the use of debate as a teaching strategy in the allied health professions. *Journal of Allied Health*, *36*(1), 13–29.

Rashtchi, M., & Sadraeimanesh, F. (2011). Is debate a useful strategy in enhancing the reading comprehension and critical thinking of Iranian EFL learners. *Theory and Practice in Language Studies*, *1*(4), 361–369. https://doi.org/10.4304/tpls.1.4.361-369

Reeves, T. C. (2006). Design research from a technology perspective. In J. van den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), *Educational design research* (pp. 52–66). Routledge.

Reinking, D., & Bradley, B. A. (2008). *On formative and design experiments: Approaches to language and literacy research*. Teachers College Press.

Reznitskaya, A., Anderson, R. C., McNurlen, B., Nguyen-Jahiel, K., Archodidou, A., & Kim, S. Y. (2001). Influence of oral discussion on written argument. *Discourse Processes*, *32*(2–3), 155–175. https://doi.org/10.1080/0163853X.2001.9651596

Richards, J. (1987). Beyond methods: Alternative approaches to instructional design in language teaching. *Prospect*, *3*(1), 11–30.

Richards, J. C., & Renandya, W. A. (Eds.). (2002). *Methodology in language teaching: An anthology of current practices*. Cambridge University Press. http://hdl.handle.net/1959.14/1087890

Richards, J. C., & Rodgers, T. S. (2001). *Approaches and methods in language teaching*. Cambridge University Press. https://doi.org/10.1017/CBO9780511667305

Romaña Correa, Y. (2015). Skype™ conference calls: A way to promote speaking skills in the teaching and learning of English. *Profile: Issues in Teachers' Professional Development*, *17*(1), 143–156. https://doi.org/10.15446/profile.v17n1.41856

Ross, J. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research and Evaluation*, *11*(1), 1–13. https://doi.org/10.7275/9wph-vv65

Rowland, R. (1995). The practical pedagogical function of academic debate. *Contemporary Argumentation and Debate*, *16*, 98–108.

Roy, A., & Macchiette, B. (2005). Debating the issues: A tool for augmenting critical thinking skills of marketing students. *Journal of Marketing Education*, *27*(3), 264–276. https://doi.org/10.1177/0273475305280533

Rubin, D. L., & Kang, O. (2008). Writing to speak: What goes on across the two-way street. In D. Belcher & A. Hirvela (Eds.), *The oral-literate connection: Perspectives on L2 speaking, writing, and other media interactions* (pp. 210–225). University of Michigan Press. https://doi.org/10.3998/mpub.231182

Rutten, K., & Soetaert, R. (2013). Rhetoric, citizenship, and cultural literacy. *CLCWeb: Comparative Literature and Culture*, *15*(3), 1–9. https://doi.org/10.7771/1481-4374.2242

Rybold, G. (2006). *Speaking, listening and understanding: Debate for non–native-English speakers*. International Debate Education Association.

Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching*, *41*(5), 513–536. https://doi.org/10.1002/tea.20009

Sampson, V., & Clark, D. B. (2008). Assessment of the ways students generate argument in science education: Current perspectives and recommendations for future directions. *Science Education*, *92*(3), 447–472. https://doi.org/10.1002/sce.20276

Sasaki, M. (2004). A multiple-data analysis of the 3.5-year development of EFL student writers. *Language Learning*, *54*(3), 525–582. https://doi.org/10.1111/j.0023-8333.2004.00264.x

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*(3), 277–301. https://doi.org/10.1075/aral.18035.sas

Sato, M., & Loewen, S. (2019a). Towards evidence-based second language pedagogy: Research proposal and pedagogical recommendations. In M. Sato & S. Loewen (Eds.), *Evidence-based second language pedagogy: A collection of instructed second language acquisition studies* (pp. 1–24). Routledge. https://doi.org/10.4324/9781351190558

Sato, M., & Loewen, S. (2019b). Do teachers care about research? The research-pedagogy dialogue. *ELT Journal*, *73*(1), 1–10. https://doi.org/10.1093/elt/ccy048

Scannapieco, F. A. (1997). Formal debate: An active learning strategy. *Journal of Dental Education*, *61*(12), 955–961. https://doi.org/10.1002/j.0022-0337.1997.61.12.tb03181.x

Schmidt, M., & Kopcha, T. (2016). Bridging theory and practice through educational design research: Two cases of technology-integration in teacher education. In G. Chamblee & L. Langub (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference* (pp. 2336–2343). Association for the Advancement of Computing in Education (AACE).

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, *14*(4), 357–385. https://doi.org/10.1017/S0272263100011189

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press. https://doi.org/10.1017/CBO9781139524780

Schulz, R. A. (1996). Focus on form in the foreign language classroom: Students' and teachers' views on error correction and the role of grammar. *Foreign Language Annals*, *29*(3), 343–364. https://doi.org/10.1111/j.1944-9720.1996.tb01247.x

Scott, S. (2008). Perception of students' learning critical thinking through debate in technology classroom: A case study. *The Journal of Technology Studies*, *34*(1), 39–44. https://doi.org/10.21061/jots.v34i1.a.5

Shamsudin, M., Othman, M., Jahedi, M., & Aralas, D. (2017). Enhancing English learners' willingness to communicate through debate and philosophy inquiry discussion. *English Language Teaching*, *10*(8), 145–152. https://doi.org/10.5539/elt.v10n8p145

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. https://doi.org/10.1080/13670050.2016.1159653

Shumin, K. (2002). Factors to consider: Developing adult EFL students' speaking abilities. In J. C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching: An anthology of current practices* (pp. 201–211). Cambridge University Press. http://hdl.handle.net/1959.14/1087890

Siebert, L. L. (2003). Student and teacher beliefs about language learning. *The Ortesol Journal*, *21*, 7–39.

Simonneaux, L. (2001). Role-play or debate to promote students' argumentation and justification on an issue in animal transgenesis. *International Journal of Science Education*, *23*(9), 903–927. https://doi.org/10.1080/09500690010016076

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics*, *17*(1), 38-62. https://doi.org/10.1093/applin/17.1.38

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press. https://doi.org/10.5070/L4111005027

Skehan, P. (2003). Task-based instruction. *Language Teaching*, *36*(1), 1–14. https://doi.org/10.1017/S026144480200188X

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and Lexis. *Applied Linguistics*, *30*(4), 510–532. https://doi.org/10.1093/applin/amp047

Skehan, P. (2014). Synthesizing and applying task research. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 211–260). John Benjamins. https://doi.org/10.1075/tblt.5

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, *49*(1), 93–120. https://doi.org/10.1111/1467-9922.00071

Smith, R. A. (1990). Are peer ratings of student debates valid? *Teaching of Psychology*, *17*(3), 188–189. https://doi.org/10.1207/s15328023top1703_13

Snider, A. (2008). *The code of the debater: Introduction to policy debating*. International Debate Education Association.

Snider, A., & Schnurer, M. (2006). *Many sides: Debate across the curriculum*. 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*(1), 67–82. https://doi.org/10.1016/j.jeap.2004.02.004

Stapleton, P. (2014). Language teaching research: Promoting a more interdisciplinary approach. *ELT Journal*, *68*(4), 432–441. https://doi.org/10.1093/elt/ccu040

Stapleton, P., & Wu, Y. A. (2015). Assessing the quality of arguments in students' persuasive writing: A case study analyzing the relationship between surface structure and substance. *Journal of English for Academic Purposes*, *17*, 12–23. https://doi.org/10.1016/j.jeap.2014.11.006

Stefani, L. A. (1994). Peer, self and tutor assessment: Relative reliabilities. *Studies in Higher Education*, *19*(1), 69–75. https://doi.org/10.1080/03075079412331382153

Stepp-Greany, J. (2002). Student perceptions on language learning in a technological environment: Implications for the new millennium. *Language Learning & Technology*, *6*(1), 165–180. https://doi.org/10125/25148

Stewart, T. (2003). Debate for ESOL students. *TESOL Journal*, *12*(1), 9–15. https://doi.org/10.1002/j.1949-3533.2003.tb00114.x

Stewart, T. (2006). Teacher-researcher collaboration or teachers' research? *TESOL Quarterly 40*(2), 421–430. https://doi.org/10.2307/40264529

Stewart, T., & Pleisch, G. (1998). Developing academic language skills and fluency through debate. *The Language Teacher, 22*(10), 27–32.

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. https://doi.org/10.1016/j.jslw.2009.02.003

Suzuki, S., & Kormos, J. (2020). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition, 42*(1), 143–167. https://doi.org/10.1017/S0272263119000421

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury.

Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *The Canadian Modern Language Review, 50*(1), 158–164. https://doi.org/10.3138/cmlr.50.1.158

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 E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–483). Lawrence Erlbaum. https://doi.org/10.4324/9781410612700

Swain, M. (2013). The inseparability of cognition and emotion in second language learning. *Language Teaching, 46*(2), 195–207. https://doi.org/10.1017/S0261444811000486

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics, 16*(3), 371–391. https://doi.org/10.1093/applin/16.3.371

Takač, V. P., & Ivezić, S. V. (2019). Frame markers and coherence in L2 argumentative essays. *Discourse and Interaction, 12*(2), 46–71. https://doi.org/10.5817/DI2019-2-46

Tannen, D. (1998). *The argument culture: Moving from debate to dialogue.* Random House.

Tanskanen, S. K. (2006). *Collaborating towards coherence: Lexical cohesion in English discourse.* John Benjamins. https://doi.org/10.1075/pbns.146

Tarone, E. (2005). Speaking in a second language. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 485–502). Lawrence Erlbaum. https://doi.org/10.4324/9781410612700

Tavakoli, P., Campbell, C., & McCormack, J. (2016). Development of speech fluency over a short period of time: Effects of pedagogic intervention. *TESOL Quarterly, 50*(2), 447–471. https://doi.org/10.1002/tesq.244

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–273). John Benjamins. https://doi.org/10.1075/lllt.11

Tessier, J. T. (2009). Classroom debate format. *College Teaching*, *57*(3), 144–152. https://doi.org/10.3200/CTCH.57.3.144-152

Thornbury, S. (2005). *How to teach speaking*. Longman.

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

Toulmin, S. (2003). *The uses of argument* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511840005

Trapp, R. (1986). The role of disagreement in interactive argument. *Journal of the American Forensic Association*, *23*(1), 23–41. https://doi.org/10.1080/00028533.1986.11951328

Tumposky, N. R. (2004). The debate debate. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *78*(2), 52–56. https://doi.org/10.3200/TCHS.78.2.52-56

Turgut, Y. (2009). EFL learners' experience of online writing by PBWiki. In G. Siemens & C. Fulford (Eds.), *Proceedings of ED-MEDIA 2009–World Conference on Educational Multimedia, Hypermedia, & Telecommunications* (pp. 3838–3847). Association for the Advancement of Computing in Education (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*(1), 36–62. https://doi.org/10.1177/0741088312469013

Van den Akker, J. (1999). Principles and methods of development research. In J. van den Akker, R. M. Branch, K. Gustafson, N. Nieveen, & T. Plomp (Eds.), *Design approaches and tools in education and training* (pp. 1–14). Kluwer.

Van den Akker, J., Gravemeijer, K., McKenney, S., & Nieveen, N. (Eds). (2006). *Educational design research*. Routledge

Van den Akker, J., & Nieveen, N. (2017). The role of teachers in design research in education. In S. Doff & R. Komoss (Eds.), *Making change happen* (pp. 75–86). Wiesbaden. https://doi.org/10.1007/978-3-658-14979-6

Van den Branden, K. (2016). Task-based language teaching. In G. Hall (Ed.), *The Routledge handbook of English language teaching* (pp. 238–251). Routledge. https://doi.org/10.4324/9781315676203

Van Eemeren, F. H., de Glopper, K., Grootendorst, R., & Oostdam, R. (2015). Student performance in identifying unexpressed premises and argumentation schemes. In F. H. van Eemeren (Ed.), *Reasonableness and effectiveness in argumentative discourse* (pp. 743–757). Springer. https://doi.org/10.1007/978-3-319-20955-5

Van Strijen, S., & Morino, A. J. (2015). *Het gebruik van de doeltaal in de lessen Spaans: de motieven van docenten en de perceptie van leerlingen* [The use of the target language in Spanish lessons: Teachers' motives and pupils' perceptions] [Master's thesis, Utrecht University, the Netherlands]. https://dspace.library.uu.nl/handle/1874/317201

Varonis, E. M., & Gass, S. (1985). Non-native/non-native conversations: A model for negotiation and meaning. *Applied Linguistics*, *6*(1), 71–90. https://doi.org/10.1093/applin/6.1.71

Venville, G. J., & Dawson, V. M. (2010). The impact of a classroom intervention on grade 10 students' argumentation skills, informal reasoning, and conceptual understanding of science. *Journal of Research in Science Teaching*, *47*(8), 952–977. https://doi.org/10.1002/tea.20358

Villarreal, I., & Gil-Sarratea, N. (2020). The effect of collaborative writing in an EFL secondary setting. *Language Teaching Research*, *24*(6) 874–897. https://doi.org/10.1177/1362168819829017

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

Wade, R. C. (1994). Teacher education students' views on class discussion: Implications for fostering critical reflection. *Teaching and Teacher Education*, *10*(2), 231–243. https://doi.org/10.1016/0742-051X(94)90015-9

Walker, A., & Kettler, T. (2020). Developing critical thinking skills in high ability adolescents: Effects of a debate and argument analysis curriculum. *Talent*, *10*(1), 21–39. https://doi.org/10.46893/talent.758473

Walker, M., & Warhurst, C. (2000). 'In most classes you sit around very quietly at a table and get lectured at …': Debates, assessment and student learning. *Teaching in Higher Education*, *5*(1), 33–49. https://doi.org/10.1080/135625100114948

Wallace, M. J. (1998). *Action research for language teachers*. Cambridge University Press.

Wallin, A., & West, E. (2013). Design and validation of teaching-learning sequences: Content-oriented theories about transmission of sound and biological evolution. In T. Plomp & N. Nieveen (Eds.), *Educational design research—Part B: Illustrative cases* (pp. 179–198). SLO.

Walton, D. N. (1996). *Argumentation schemes for presumptive reasoning*. Lawrence Erlbaum Associates.

Wang, J. (2016). On freeman's argument structure approach. *CLAR*, 23–29.

Warner, E., & Bruschke, D. (2001). *"Gone on debating:" Competitive academic debate as a tool of empowerment for urban America* [Paper presentation]. Western States Communication Association Convention, Coer d'Alene, Idaho.

Weaver, C. (2007). Willingness to communicate: A mediating factor in the interaction between learners and tasks. In K. van den Branden, K. van Gorp, & M. Verhelst (Eds.), *Tasks in action: Task-based language education from a classroom-based perspective* (pp. 159–193). Cambridge Scholars Publishing.

Weinberg, L. (2017). Motivation, technology and language learning. In R. Breeze & C. Sancho Guinda (Eds.), *Essential competencies for English-medium university teaching* (pp. 295–307). Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-40956-6

Weir, R. (2009). Can we discuss this? *Inside higher education*. Retrieved February 10, 2016, from https://www.insidehighered.com/advice/2009/09/09/can-we-discuss

Weissberg, R. (2006). *Connecting speaking & writing in second language writing instruction*. University of Michigan Press. https://doi.org/10.3998/mpub.6702

West, L., & Verspoor, M. (2016). An impression of foreign language teaching approaches in the Netherlands. *Levende Talen Tijdschrift*, *17*(4), 26–36.

Widdowson, H. G. (1998). Context, community, and authentic language. *TESOL Quarterly*, *32*(4), 705–716. https://doi.org/10.2307/3588001

Williams, C. (2006). Providing challenge and engagement in classroom learning for G&T students. Retrieved June 26, 2014, from https://giftedphoenix.files.wordpress.com/2012/11/nagty-providing-challenge-and-engagement-in-classroom-learning.pdf

Williams, D. E., McGee, B. R., & Worth, D. S. (2001). University student perceptions of the efficacy of debate participation: An empirical investigation. *Argumentation and Advocacy*, *37(4)*, 198-209. https://doi.org/10.1080/00028533.2001.11951670

Williams, J. (2008). The speaking-writing connection in second language and academic literacy development. In D. Belcher & A. Hirvela (Eds.), *The oral-literate connection: Perspectives on L2 speaking, writing, and other media interactions* (pp. 10–25). University of Michigan Press. https://doi.org/10.3998/mpub.231182

Williams, J. (2012). The potential role(s) of writing in second language development. *Journal of Second Language Writing*, *21*, 321–331. https://doi.org/10.1016/j.jslw.2012.09.007

Willingham, D. T. (2007). Critical thinking: Why is it so hard to teach? *American Educator*, *31(2)*, 8–19. https://doi.org/10.3200/AEPR.109.4.21-32

Willis, J. (1992). Inner and outer: Spoken discourse in the language classroom. In M. Coulthard (Ed.), *Advances in spoken discourse analysis* (pp. 162–182). Routledge. https://doi.org/10.4324/9780203200063

Willis, J. (1996). *A framework for task-based learning*. Longman.

Willis, J. (1998). Task-based learning: What kind of adventure. *The Language Teacher*, *22*(7), 17–18.

Wingate, U. (2012). 'Argument!' helping students understand what essay writing is about. *Journal of English for Academic Purposes*, *11*(2), 145–154. https://doi.org/10.1016/j.jeap.2011.11.001

Wolfe–Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. University of Hawai'i Press.

Wood, S., & Rowland-Morin, P. (1989). Motivational tension: Winning vs. pedagogy in academic debate. *National Forensics Journal*, 7, 81–98.

Worthen, T. K., & Pack, G. N. (1992). *Classroom debate as an experiential activity across the curriculum* [Paper presentation]. The Annual Meeting of the Speech Communication Association, Chicago, USA.

Woude, N. N. V., Janssen, D., & Sanders, T. (2011). Flywheel forensics: Debat en public speaking in de klas [Flywheel forensics: Debate and public speaking in class]. *Levende Talen Magazine*, 98 *(8)*, 5–9.

Woude, N. V. D., Janssen, D., & Sanders, T. (2012). Hear, hear! Leer scholieren debatteren [Hear, hear! Teach students how to debate]. *Levende Talen Magazine*, 99*(5)*, 10–13.

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*(3), 601–623. https://doi.org/10.17239/jowr-2021.12.03.03

Xargia, M. (2016). A comparative analysis of Greek vs. English: Can it be argued that teenagers can argue? *Selected Papers on Theoretical and Applied Linguistics*, *21*, 839–854. https://doi.org/10.26262/istal.v21i0.5273

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. https://doi.org/10.3998/mpub.231182

Yang, Y. (2016). Teaching Chinese college ESL writing: A genre-based approach. *English Language Teaching*, *9*(9), 36–44. https://doi.org/10.5539/elt.v9n9p36

Yasuda, S. (2019). Children's meaning-making choices in EFL writing: The use of cohesive devices and interpersonal resources. *System, 85*, 1–15. https://doi.org/10.1016/j.system.2019.102108

Yeh, S. S. (1998). Empowering education: Teaching argumentative writing to cultural minority middle-school students. *Research in the Teaching of English*, *33*(1), 49–83. https://www.jstor.org/stable/40171572

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. http://hdl.handle.net/1959.3/82391

Yoon, H. J. (2017). Linguistic complexity in L2 writing revisited: Issues of topic, proficiency, and construct multidimensionality. *System*, *66*, 130–141. https://doi.org/10.1016/j.system.2017.03.007

Yoon, H. J., & Polio, C. (2017). The linguistic development of students of English as a second language in two written genres. *TESOL Quarterly*, *51*(2), 275–301. https://doi.org/10.1002/tesq.296

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, *24*(1), 1–27. https://doi.org/10.1093/applin/24.1.1

Zare, P., & Othman, M. (2013). Classroom debate as a systematic teaching/learning approach. *World Applied Science Journal*, *28*(11), 1506–1513. https://doi.org/10.5829/idosi.wasj.2013.28.11.1809

Zare, P., & Othman, M. (2015). Students' perceptions toward using classroom debate to develop critical thinking and oral communication ability. *Asian Social Science*, *11*(9), 158–170. http://dx.doi.org/10.5539/ass.v11n9p158

Zorwick, L. W., & Wade, J. M. (2016). Enhancing civic education through the use of assigned advocacy, argumentation, and debate across the curriculum. *Communication Education*, *65*(4), 434–444. https://doi.org/10.1080/03634523.2016.1203005

Appendix A

Debate preparation form

Name:_____

Issue:_____

Mention three arguments that support your side:

1. ……...
2. ……...
3. ……...

Mention three arguments the opposing team is likely to use:

1. ……...
2. ……...
3. ……...

List at least **three** new words that are relevant for the debate (and that you are going to use):

……
……
……

State and justify your opinion using +/- 80 words. (**Use your own words**)

……
……
……
……
……
……
……
……
……
……
……
……
……
……
……
……
……
……
……
……

AP

Mention the sources that you used in your research. Attach them to this document and **indicate** where the information relevant to this issue is discussed in the text:

……
……
……

During debate

Points for your rebuttals:

……
……
……
……
……
……
……
……
……
……
……
……
……
……
……
……

Mistakes of your classmates. Write at least three:

……
……
……
……
……
……
……

Words being learned during the debate. Write at least three:

……
……
……
……
……
……
……

Appendix B

The design principles underlying the debate task design

Stages and activities	Design principles	Rationale (in order to …)
	Macrodesign principles	
	D.P.1 Granting instructors the freedom to operationalize tasks in a way that fits well with their teaching perspective and context	Tailor the intervention to instructors' teaching style and perspective
	D.P.2 Attending to the factors that affect students' motivation in task performance	Increase students' motivation to try their best during task completion
	Microdesign principles	
Preparatory stage		
PowerPoint presentation	D.P.3 Providing a model	Inform students of the debate project and acquaint them with the mechanics of debate
Selecting topics	D.P.2a Facilitating the selection of interesting and engaging debate topics	Select interesting and engaging debate topics
Pre-debate stage		
Reading articles and writing cases	D.P.4 Employing nontask preparation activities D.P.4a Requiring students to read articles and summarize them during the preparation stage D.P.4b Requiring students to write cases in which they should use cohesive markers and the new learned words (when possible) during the preparation stage	Boost performance in the during-task phase Enable students to become knowledgeable about the debate topic and search for relevant arguments Enable students to produce coherent constructive speeches
Rehearsing cases	D.P.5 Facilitating task repetition	Boost speech delivery
During-debate stage		
Pre-task planning	D.P.6 Encouraging and facilitating pre-task planning	Boost performance in the during-task phase
Debate formats	D.P.9 Selecting debate formats that have a clear structure and that ensure equal involvement of all students	Facilitate smooth and engaging debates
Focus on form	D.P.7 Implementing focus on form through corrective feedback	Attend to students' accuracy
Grading debates	D.P.10 Grading debates to prompt students to take them more seriously	Stimulate students to take debates seriously
Post-debate stage		
Focus on forms	D.P.8 Employing focus on forms activities in post-task phase	Attend to students' accuracy

Appendix C

Questionnaire about students' attitude towards debate

The questionnaire consists of several statements. We would like you to indicate for each statement the extent to which you agree with it. There are no right or wrong answers; it's all about your opinion. After you have indicated the extent to which you agree with each statement, you may also give comments. The questionnaire is anonymous, so feel free to give your opinion honestly. Thank you very much!

	Strongly disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly agree 5
1. Debating is interesting.					
2. I like debating because it offers challenges.					
3. Debating is fun because it is more appealing than working with course books.					
4. I will always participate in debating in class.					
5. I would like my teacher to schedule debating in class often.					
6. Debating is pointless.					
7. Because of debating I am actively involved in the lesson and I like that.					
8. I like debating because it requires me to play an active role in class.					
9. Debating is amusing.					
10. Rebuttal is a fun challenge.					
11. I like debating because it makes me think.					
12. Debating is important for my listening skills and that's why I like it.					
13. Defending a point of view is a fun challenge.					
14. Debating is entertaining.					
15. Debating pushes me to be active and I like that.					
16. I would rather debate than work with course books.					
17. I like debating because it allows me to work together with my classmates.					

	Strongly disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly agree 5
18. Debating is boring.					
19. Debating is important for my speaking skills and that's why I like it.					
20. Debating in a group is not fun.					
21. During debating I have to think critically and I don't like that.					
22. Because of debating I am more anxious about speaking in English.					
23. Debating challenges me and I like that.					
24. Debating is fun because it pushes me to think.					
25. Debating is fun because it allows me to be active in class.					
26. Because of debating I don't have to work with course books and that's why I like it.					
27. Convincing my classmates is a challenge that I enjoy.					
28. I like debating because it stimulates me to think critically.					
29. I like debating because it increases my vocabulary.					
30. Debating requires me to play a passive role in class.					
31. Debating is fun.					
32. Debating requires me to think deeply and I like that.					
33. Debating is fun because it gives me the opportunity to work in a group.					
34. Debating in a group with my classmates is fun.					
35. Debating allows me to work with my classmates and I like that.					
36. Debating is important for my writing skills and that's why I like it.					

37. I find debating more enjoyable than working with course books.					
38. Debating is pleasant.					
39. I would rather work with course books than debate.					
40. I like debating.					

Comments

……
……
……
……
……
……
……
……
……
……
……
……

Please fill in the following information

Gender:

- ○ Male
- ○ Female

Age: …….

This is the end. Thank you for filling in this questionnaire!

Appendix D

Questionnaire about students' perceptions of the effects of debate

You have participated in a number of debate activities. To facilitate getting your opinion about these activities, we have developed this questionnaire. With debating, we mean the whole activity, including preparation (e.g., reading articles). The questionnaire consists of several statements. You need to indicate for each statement the extent to which you agree with it. There are no right or wrong answers.

After completing the questionnaire, you can also give comments about the debate activities. Finally, the questionnaire is anonymous, so feel free to express your opinion honestly. Thank you!

Part 1

	Strongly disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly agree 5
1. Debating is instructive.					
2. My listening proficiency in English has improved as a result of debate activities.					
3. My speaking proficiency in English has improved as a result of debate activities.					
4. My writing proficiency in English has improved as a result of debate activities (by means of preparation and making notes during debates).					
5. Debate activities have not brought any improvement to my speaking proficiency in English.					
6. Debate activities have not brought any improvement to my grammar in English.					
7. Debate activities have not brought any improvement to my writing proficiency in English.					
8. Debating is boring.					
9. My reading proficiency in English has improved as a result of debate activities (by means of preparation).					
10. Because of debate activities I now make fewer grammatical mistakes in English.					
11. My vocabulary in English has improved as a result of debate activities.					
12. Debating is an important part of the English curriculum.					

13. Debate activities have not brought any improvement to my listening proficiency in English.					
14. Debate activities have not brought any improvement to my reading proficiency in English.					
15. I have not learned any new English words as a result of debate activities.					

Part 2

Have you got the impression that …	**Definitely** **1**	**2**	**3**	**4**	**Definitely Not** **5**
16. you can understand TV shows in English better as a result of debate activities?					
17. you can now speak better English with other people as a result of debate activities?					
18. you can now write better English as a result of debate activities?					
19. you can now read and understand English texts more easily as a result of debate activities?					
20. you have learned new English words as a result of debate activities?					
21. you now make fewer grammatical mistakes in English as a result of debate activities?					

Comments

……
……
……
……
……
……
……
……
……
……

Please fill in the following information

Gender:

- ○ Male
- ○ Female

Age: …….

This is the end. Thank you for filling in this questionnaire!

Appendix E

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 weren't 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	10.22
	C/T	1.80
Lexical complexity	MLTD	95
	Word frequency	2.99
	Word length	4.14
Accuracy	Error-free clauses	0.84
	Lexical errors per 100 words	0
	Syntactic errors per 100 words	0
	Morphological errors per 100 words	2.17
	Preposition errors per 100 words	0
Cohesion	Transition markers	3
	Frame markers	2
	Gloss markers	0
	Conclusion markers	0
	Diversity type	2
	Diversity 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.

Appendix F

Framework for the analysis of argumentation structure

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

Appendix G

Scale for the assessment of the overall structural complexity of argumentation

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

Appendix H

Framework for the analysis of argumentation quality

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

Appendix I

Examples of scoring procedures for two written texts

Example 1

I am against legalizing abortion, because of the following reasons. **Claim**
First of all, being able to bear children is a gift of God and at the moment you decide to kill it, you're deciding against the will of God. **Data 1**
Second, abortion could lead to mental and physical pain. **Data 2**
An example of mental pain is that a lady could get suicidal tendencies. **Subordinate argument 1 (data 2)**
An example of the physical pain, is the pain while the operation is finding place. **Subordinate argument 2 (data 2)**
Third, while you're utilizing abortion, you are murdering a child. **Data 3**
A baby is a human being and has also rights. **Data 4**
Last of all, there are ladies with the wish to bear children. It is not fair against those ladies which are not even able to bear children. **Data 5**

Argumentation structural analysis	**Assessment**
Overall argument complexity	2
Claim	1
Data	5
Subarguments	2
Quality of arguments analysis	
Overall assessment	3
Organization of arguments	3.5
Sufficiency of arguments	3
Clarity of arguments/ comprehensibility	3
Elaboration of arguments	3
Relevance of arguments	3.5
Persuasiveness of arguments	3
Addressing the opposing view	1

Example 2

I believe that abortion should stay legal. These are the arguments which I base my opinion on. **Claim**
First and foremost, women should do whatever they want with their body. You should not be the one who makes the decisions for her. **Data 1**
Furthermore, there are sometimes babies born which are unwanted. **Data 2**
Mothers will not care for their baby as they should. **Subordinate argument 1 (data 2)**
There are also personal issues involving abortion. Many mothers don't want to raise the baby because they want to have a successful career or finish their school. **Data 3**
Moreover, there are several diseases which can be transferred through sexual interaction. **Data 4**
You don't want a baby coming into this world with aids or any other disease. **Subordinate argument 2 (data 4)**
Besides, in 1970 abortion was illegal. This meant that mothers who do not wish to bring a baby into this world, had to perform unsafe abortion in secrecy. As a result, there were a lot of mothers who died after the abortion. **Data 5**
In today's world, you will be able to go through an abortion, and you will have no after effects from the surgery. **Data 6**
Also, mothers who are addicted to drugs, alcohol or smoking will not prioritize the health of their baby. **Data 7**
Lastly, people say that adoption is an option for abortion. **Counterargument**
If there was no abortion performed, there would be no place for many babies in foster homes or abortion babies, because there isn't enough place. **Rebuttal**

Argumentation structural analysis	**Assessment**
Overall argument complexity	4
Claim	1
Data	7
Subarguments	2
Counterarguments	1
Rebuttals	1
Quality of arguments analysis	
Overall assessment	4
Organization of arguments	4
Sufficiency of arguments	4
Clarity of arguments/ comprehensibility	4
Elaboration of arguments	3.5
Relevance of arguments	4
Persuasiveness of arguments	4
Addressing the opposing view	3

Summary in Dutch

Debatteren wordt door verschillende onderzoekers gezien als een effectief instrument voor taalvaardigheidsonderwijs. Maar vooralsnog is er weinig empirisch onderzoek verricht waarin die effecten daadwerkelijk worden aangetoond. Met name onderzoek naar debatteren in onderwijs in een vreemde taal ontbreekt. Dit onderzoek beoogt hierin verandering te brengen. In dit proefschrift wordt nagegaan in hoeverre de Engelse taal- en argumentatievaardigheid van Nederlandse middelbare scholieren profiteert van debatteren in de klas.

Achtergrond van de studie

Voor onderwijs in Engels als de tweede taal (L2) bestaat de behoefte aan didactisch innovatieve leermiddelen om het leren aantrekkelijker en effectiever te maken voor leerlingen. Debatdidactiek lijkt in theorie goede papieren te hebben om de problemen bij het onderwijzen en leren van L2 in middelbare scholen het hoofd te bieden (zie hoofdstuk 3). Om de effectiviteit van debatteren in de klas empirisch te onderzoeken zijn verschillende experimenten opgezet en uitgevoerd op drie middelbare scholen. Daarin is nagegaan welke effecten debatdidactiek heeft op de spreek- en schrijfvaardigheid en de argumentatievaardigheid in het Engels.

Hoofdstuk 1 presenteert de onderzoekscontext en beredeneert van daaruit het belang van het onderzoek. Hoofdstuk 2 gaat dieper in op de literatuur over debatteren als onderwijsinstrument. In dit hoofdstuk worden een aantal vaak genoemde voordelen besproken van debatteren in de klas. Het accent daarbij ligt op de verbetering van taal- en argumentatievaardigheden omdat deze ook centraal staan in dit onderzoek.

De literatuurstudie van hoofdstuk 2 bracht aan het licht dat bestaand instructiemateriaal en bestaande debatmodellen voor ons doel ongeschikt waren. Ze waren vooral ontwikkeld voor competitieve debatten in de moedertaal. Om de effecten van debatteren op taal- en argumentatievaardigheden in de L2 middelbare schoolcontext te onderzoeken, was het daarom nodig om een nieuw debattaakontwerp te ontwikkelen, gericht op betekenisvolle uitwisseling van argumenten, en dat leerlingen helpt bij de voorbereiding en deelname aan debatten in de klas. In hoofdstuk 2 bespreken we de randvoorwaarden voor de constructie van een debattaakontwerp dat in hoofdstuk 3 verder is uitgewerkt en getoetst.

Debattaakontwerp

Hoofdstuk 3 beschrijft de ontwikkeling van een debattaakontwerp waarin rekening is gehouden met de specifieke kenmerken van de L2 context in het Nederlandse voortgezet onderwijs. Hierin zijn de richtlijnen van *educational design research* gevolgd (EDR) en wel in drie fasen. In de

eerste fase is een contextanalyse uitgevoerd met als doel de onderwijskundige problemen en de didactische mogelijkheden van debatteren in de klas in kaart te brengen. Vervolgens is op basis van een literatuurstudie het conceptuele kader van de studie opgesteld waarmee het eerste prototype van de studie kon worden ontwikkeld dat nader is verfijnd op basis van klassenobservaties, groepsdiscussies, vragenlijsten, en interviews. Deze debattaak is verder in het hele vervolgonderzoek ingezet om de verschillende onderzoeksvragen te beantwoorden.

Attitudes ten opzichte van debatteren

In hoofdstuk 4 wordt nagegaan in hoeverre L2-leerlingen plezier beleven aan debatteren in de klas en waarom. Uit het onderzoek blijkt dat dat de deelnemers een sterk positieve houding hebben tegenover debatteren. Kenmerken die hierbij genoemd worden zijn: *plezier*, *actieve deelname*, *kritisch denken*, *uitdaging*, *samenwerken*, *taalvaardigheid*, en *debat versus lesboek*.

In deze studie is ook nagegaan of er verschillen zijn tussen jongens en meisjes in de perceptie van debatteren in de klas. Daaruit bleek dat meisjes meer waarde hechten aan aspecten als *plezier*, *actieve deelname*, en *uitdaging*, en jongens meer aan *kritisch nadenken*.

Percepties ten opzichte van debatteren

Hoewel verschillende studies het potentieel van debatteren in de klas hebben onderkend, is er tot nog toe weinig onderzoek verricht dat effecten op L2 taalontwikkeling ook daadwerkelijk heeft aangetoond. Ook is er weinig bekend over de mate waarin leerlingen zelf vinden dat deze vorm van vreemde-taalonderwijs leerzaam is. In hoofdstuk 5 is daarom nagegaan welke effecten van debatteren leerlingen bij zichzelf waarnemen; in hoeverre zij zelf denken dat hun spreek, schrijf-, lees- en luistervaardigheden, woordenschat, en grammatica er baat bij hebben.

Surveys, interviews, en debatevaluaties lieten zien dat leerlingen sterk de indruk hebben dat ze door het debatteren beter gaan spreken, schrijven, lezen, en luisteren in de vreemde taal. Ook blijken oordelen over de effecten van het debatteren op de eigen vaardigheden sterk samen te hangen met hun cijfers voor debatteren en het gemiddelde cijfer voor Engels.

Effect van debatdidactiek op de spreekvaardigheid

Hoewel debatteren in de klas vaak wordt beschreven als een potentieel effectief middel om de spreekvaardigheid van leerlingen te verbeteren, is er geen onderzoek bekend waarin dat verband ook daadwerkelijk is aangetoond. In hoofdstuk 6 is daarom nagegaan wat de effecten zijn van de debatdidactiek op de ontwikkeling van de spreekvaardigheid in het Engels. Voor dit deelonderzoek is gebruik gemaakt van een pretest/posttest-design met een controlegroep. Tijdens de voor- en nameting hebben leerlingen mondelinge opinietaken uitgevoerd, die zijn geanalyseerd op verschillende aspecten van de spreekvaardigheid, zoals vlotheid, coherentie, en

correctheid. De resultaten lieten zien dat de interventiegroep na debatonderwijs beter presteerde dan de controlegroep op de meest onderzochte variabelen. Na de interventie produceerden de leerlingen langere bijdragen, spraken ze in een hoger tempo met meer coherentie, met minder fouten en meer lexicale complexiteit dan de leerlingen in de controlegroep.

Effect van debatdidactiek op schrijfopinietaken

Wanneer men spreekt over "debatteren" gaat de eerste gedachte doorgaans uit naar mondelinge communicatie en argumentatievaardigheden. Maar bij — onder meer — de voorbereiding van debatteren worden vaak ook schrijftaken uitgevoerd, en in die zin kan debateronderwijs ook bijdragen aan de ontwikkeling van de schrijfvaardigheid in L2. In een gecontroleerd pre-/posttest experiment met een controlegroep zijn we nagegaan of die effecten ook empirisch zijn aan te tonen. Daarvoor schreven leerlingen een korte opinietaak waarin zij een standpunt moesten verdedigen voor of tegen een controversieel beleidsissue. De geproduceerde teksten zijn vervolgens op verschillende dimensies geanalyseerd, bijvoorbeeld op: coherentie, cohesie, taalfouten, syntactische en lexicale complexiteit. Multilevel analyses lieten zien dat de leerlingen in de interventiegroep significant verbeterden op een groot aantal dimensies in vergelijking met de controlegroep. Ten opzichte van de leerlingen in de controlegroep schreven de leerlingen in de interventiegroep significant langere teksten die meer syntactische en lexicale complexiteit vertoonden, minder taalfouten bevatten, en een meer verfijnd en breder scala van coherentiemarkeerders etaleerden.

Effect van debatdidactiek op argumentatievaardigheid

In hoofdstuk 8 is nagegaan welke effecten de L2-debatdidactiek heeft op argumentatievaardigheden. Daarbij hebben we zowel aandacht besteed aan structurele kenmerken van redeneringen (bijvoorbeeld sub-argumenten en weerleggingen) als aan de overtuigingskracht van argumentatie (bijvoorbeeld uitwerking en relevantie van argumenten) in zowel schriftelijke als mondelinge argumentatie. In dit experiment hebben we gebruik gemaakt van een pretest/posttestdesign met een controlegroep. De resultaten lieten opnieuw positieve effecten zien. Na de interventie waren de debaters in vergelijking met de controlegroep in staat om hun standpunten beter te onderbouwen met meer argumenten en meer complexe argumentatiestructuren. Ze gebruikten niet alleen meer argumenten, maar ook meer ondersteunende argumenten (*subarguments* en *backings*). Ook kregen de verbanden tussen argumenten en conclusies meer aandacht (door middel van *warrants*) en waren ze zich meer bewust van de overtuigingskracht van hun argumentatie blijkens het gebruik van *qualifiers*. Ten slotte besteedden de leerlingen na de debatinterventie meer aandacht aan tegenargumentatie via *rebuttals*. Na de posttest slaagden de interventiegroep ook erin om overtuigender te redeneren. Bij de

meest onderzochte variabelen is er geen verschil te zien tussen de prestaties van leerlingen op mondelinge en de schriftelijke opinietaken.

Effect van debatdidactiek op het schrijven van argumentatieve essays

In hoofdstuk 9 wordt de potentie van debatdidactiek onderzocht voor het verbeteren van de kwaliteit van argumentatieve essays. Het doel hiervan was om na te gaan of de verbeterde schrijfprestaties op korte teksten zich zou vertalen in betere prestaties op een uitgebreidere schrijftaak.

Om de effecten van debatdidactiek op het schrijven van argumentatieve essays te meten, analyseerden we de argumentatieve essays van de deelnemers (in de interventie- en controlegroepen) op drie tijdstippen (voor-, na-, en uitgestelde nameting) op verschillende tekstuele en inhoudelijke kenmerken. De resultaten toonden aan dat de interventiegroep significant beter presteerde dan de controlegroep op de meeste dimensies. Na de interventie en tijdens de uitgestelde nameting schreef de interventiegroep langere, meer coherente en meer overtuigende essays met complexere syntactische structuren en minder taalfouten.

Conclusies en aanbevelingen

Debatdidactiek blijkt goede papieren te hebben om de taal- en argumentatievaardigheid van L2-leerlingen te helpen verbeteren. Daar zijn een aantal verklaringen voor te geven: het competitie-element, de positieve attitude van leerlingen, en het feit dat leerlingen hun taalgebruik oefenen voor een realistisch en aanwezig publiek. Deze eigenschappen noemen we de *macro-didactische kenmerken*. Daarnaast is de debatdidactiek gericht op samenwerken, interactie en betrokkenheid, en daarmee in het bijzonder studentgericht. Deze eigenschappen noemen we de *micro-didactische kenmerken*. Wij menen dat de gevonden resultaten te danken zijn aan de synergetische bijdrage en interactie van een groot aantal van deze factoren.

Debatteren biedt unieke leermogelijkheden en verdient daarmee een plaats in curricula voor de vreemde talen. Tijdens het debatteren en bij de voorbereiding van debatten voeren leerlingen zinvolle en betekenisvolle taken uit die vruchten afwerpen. De effecten daarvan zijn in deze studie onderzocht voor het Engels, maar er is geen enkele reden om aan te nemen dat de effecten anders zouden zijn in andere vreemde talen. Het voorgestelde debattaakontwerp (Hoofdstuk 3) leent zich gemakkelijk voor aanpassing aan meerdere omgevingen, ook buiten de Nederlandse onderwijscontext.

List of Publications

El Majidi, A., de Graaff, R., & Janssen, D. (2015a). Invest in what energizes students to learn: Investigating students' attitude towards debate in the foreign language classroom. *Journal of Language Teaching and Research, 6*(5), 924–932. https://doi.org/10.17507/jltr.0605.03

El Majidi, A., de Graaff, R., & Janssen, D. (2015b). Debatteren in het Engels. Leerzame ingrediënten in de ogen van leerlingen [Debating in English. Instructive ingredients in the eyes of students]. *Levende Talen Tijdschrift, 16*(1), 24–33.

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*(1), 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*(4), 804–821. https://doi.org/10.1111/modl.12673

El Majidi, A., Janssen, D., & de Graaff, R. (2021a). The effects of in-class debates on argumentation skills in second language education. *System, 101*, 1–15. https://doi.org/10.1016/j.system.2021.102576

El Majidi, A., Janssen, D., & de Graaff, R. (2021b). Het effect van debatdidactiek op schrijven in een vreemde taal: Een experimenteel onderzoek naar de effecten van debatonderwijs op de Engelse schrijfvaardigheid [The effect of debate pedagogy on writing in a foreign language: An experimental study of the effects of debate instruction on English writing skills]. *Levende Talen Tijdschrift, 22*(4), 3–14.

El Majidi, A., de Graaff, R., & Janssen, D. (in press). Debate as a pedagogical tool for developing speaking skills in second language education. *Language Teaching Research*. https://doi.org/10.1177/13621688211050619

El Majidi, A., de Graaff, R., & Janssen, D. (submitted). Debate pedagogy as a conducive environment for L2 argumentative essay writing.

El Majidi, A., Janssen, D., & de Graaff, R. (submitted). Het effect van debatteren op de structurele complexiteit van argumentatievaardigheden in een vreemde taal [The effect of debating on the structural complexity of argumentation skills in a foreign language].

Acknowledgments

All praise be to God, the Almighty, the Greatest of all, for giving me the perseverance, determination, patience, well-being, and strength to arrive at this point in life. Finishing this thesis has left with me mixed feelings. On the one hand, I am happy to have successfully completed this journey, but on the other hand, it is a poignant moment to come to the end of a research project that I have worked on for several years.

Completing this doctoral thesis journey without the unswerving support of my supervisors, family, and friends would have been a next-to-impossible mission. Therefore, I wish to register my heartfelt gratitude to a number of people who have played an important role in bringing this research project to a successful conclusion.

First and foremost, I wish to extend my sincere gratitude to my supervisors, prof. dr. Rick de Graaff and dr. Daniel Janssen, without whom this thesis would not have been possible. From start to finish, they have continuously provided me with valuable guidance, advice, encouragement, and countless insightful comments and suggestions that have improved the thesis significantly. Although there is much I have yet to learn about the world of research, I truly believe that I am now a better researcher, thinker, and writer because of them. Rick and Daniel, it was an honor to work with you!

Throughout my doctoral journey, I have also been blessed with help and support from many other people. Words cannot express how grateful I am for their help and generosity. I would like to express my appreciation to those who helped me with statistical assistance without which I would probably still be struggling to figure out how to analyze my data. Special thanks are also due to those who shared with me their knowledge of how to analyze L2 fluency. I also wish to express my greatest appreciation to the two teacher participants in this project Mrs. Anne van der Geest-Schoenmakers M.A. and Mrs. Madelon Elise Wolters M.A. and their schools. They participated in my research despite their very heavy workloads. My gratitude also goes to dr. Abdelbasset Dahraoui for his generous assistance with coding a large amount of data. He was always there willing to do this arduous work, which he performed conscientiously.

My family equally played an indispensable role in the completion of this project. In particular, I owe the achievement of completing this thesis to my wife, Noura. Words are insufficient to express my gratitude for her patience, understanding, sacrifice, support, and encouragement, all of which made this thesis possible. She will always be a part of any success I achieve in life. I

am also blessed with my beloved children, Assia, Maryam, Ilyas, and Adam, who bore with me throughout these busy years and are extremely proud of my accomplishment.

I would like to express a special note of gratitude to my father and mother (who, alas, is not here to witness this achievement), and also to my brothers, sisters, other family members, and friends for their continued support over the last few years. I wish my brother Brahim and my sister Zinab good luck with their PhDs. My joy will be complete when you also get successfully to the finishing line.

I am also grateful to my school for the support I received in this journey and to DUDOC-ALFA for the opportunity to participate in their program, which was truly rewarding. Last but not least, special thanks go to the Dutch Ministry of Education, Culture and Science and NWO for providing financial support to allow me to conduct this research. I hope they continue with this initiative to give more teachers the opportunity to carry out PhD research because the potential payoff is large enough to justify the effort and means invested.

About the Author

The author of this dissertation was born in Driouch, Morocco, on 20th March 1980. After receiving his bachelor's degree in English language and literature from Mohammed I University (Oujda, Morocco) in 2002, he moved to the Netherlands. In 2004, he earned another bachelor's degree in English language and culture from the University of Amsterdam where he completed a master's degree in general linguistics in 2005. After obtaining his teaching degree from Utrecht University in 2006, he started working as an English teacher at a secondary school in Rotterdam where he also prepares students for the Cambridge and Business English exams. His interest in second language pedagogy and the processes underlying second language acquisition drove him to submit a research proposal to NWO (Nederlandse Organisatie voor Wetenschappelijk Onderzoek/ Dutch Research Council). In 2017, he was awarded a five-year research grant and was accordingly accepted as a PhD candidate at Utrecht University while continuing to teach English at Montfort secondary school. This grant gave him the chance to investigate the effects of the pedagogical potential of debate in English for improving speaking, writing, and argumentation skills. He has presented his research findings in different journals and at several (inter)national conferences and educational institutions.

Debate is widely hailed as an effective teaching tool capable of attaining many learning objectives. Yet, in second language education there is scant empirical evidence available to support this assumption. The PhD study set out to: 1) elicit L2 students' attitudes towards debate and their perceptions of its impact on different aspects of language development; and 2) experimentally explore the affordances and effectiveness of debate pedagogy with regard to speaking, writing, and argumentation skills in the Dutch secondary school context.

Drawing on data from students' performance, questionnaires, interviews, and classroom observations, the main finding was that the participants greatly admire debate and that debate pedagogy can improve a number of aspects of the debaters' speaking, writing, and argumentation skills. From the findings, it can be concluded that debate pedagogy can provide a motivating learning environment with a unique pedagogical configuration that is conducive to L2 language and argumentation performance.

Debates are motivational, engaging, meaningful, challenging, interactive, student-centered, audience-oriented, and instructive. Instructors need to recognize the potential of debates and systematically employ them in their teaching practice.

ISBN: 978-94-6458-190-4