

The Effects of an EFL and L2 Russian Teletandem Class: Student Perceptions of Oral Proficiency Gains

Jennifer Bown

Brigham Young University

Correspondence regarding this article should be addressed to Jennifer Bown, Department of German and Russian, Brigham Young University, 3117 JFSB, Provo, Utah, USA. E-mail: jennifer_bown@byu.edu

Laura Catharine Smith

Brigham Young University

Correspondence regarding this article should be addressed to Laura Catharine Smith, Department of German and Russian, BYU, 3117 JFSB, Provo, Utah, USA. E-mail: laurasmith@byu.edu

Ekaterina V. Talalakina

National Research University Higher School of Economics

Correspondence regarding this article should be addressed to Ekaterina V. Talalakina, Department of Foreign Languages, HSE, Malaya Ordynka, 17/1, Moscow, Russia. E-mail: etalalakina@hse.ru

In response to the growing demand for highly proficient foreign language (L2) speakers in professional work settings, scholars and educators have increasingly turned their attention to methods for developing greater fluency in their learners who aspire to such jobs. Engaging in persuasive writing and argumentation has been shown to promote both written and oral proficiency among advanced L2 learners (Brown, 2009). This study focuses on the application of the American Council on the Teaching of Foreign Languages (ACTFL) proficiency guidelines and standards to the design of teletandem courses in English as a Foreign Language (EFL) and Russian as a Foreign Language developed to promote Advanced and Superior-level language gains. ACTFL Can-Do statements were used to evaluate learners' self-reported language gains as a result of participating in the course. The results indicated that such an approach can indeed yield significant perceived gains, especially for spoken language, for all the participants regardless of their target language and home institution.

Keywords: Proficiency, ACTFL Standards, English, Russian, teletandem

Introduction

As the world becomes ever more globalised, the demand for global professionals competent in both foreign languages (L2) and cultures continues to grow as well. Consequently, the newly globalised economy requires even higher levels of foreign language proficiency. Martin (2015) notes that more and more employers require learners with Advanced and Superior-levels of proficiency according to the scale established by the American Council on the Teaching of Foreign Languages (ACTFL, cf. also Swender, 2012), the descriptions of which are outlined below. This demand for higher levels of proficiency has created a critical challenge for many language programs where it is not uncommon for some graduating language majors to fall short of this proficiency threshold. With a shift in recent years towards teaching for proficiency, scholars have increasingly turned their attention to better facilitating the development of these professional levels of language proficiency, moving students towards Advanced (and to some extent Superior) levels of proficiency. To this end, at least three volumes dedicated to promoting advanced levels of proficiency have appeared in print over the past two decades

(Brown, & Bown, 2015; Leaver & Shekman, 2002; Murphy & Evans-Romaine, 2017) in addition to other studies in journals (e.g., Darhower, 2014; Donato & Brooks, 2004).

One method of instruction that has proven effective for promoting high levels of proficiency is the use of debate and argumentation (Brown, 2009; Brown, Talalakina, Yakusheva, & Eggett, 2012). As Brown (2009) points out, the task of defending and supporting opinions is a core function of the Superior level, and the criteria for Superior as outlined in the ACTFL proficiency guidelines dovetail with functions emphasized in public speaking and debate. Moreover, using the L2 as a medium for debate represents a form of task-based instruction, where learners are focused on conveying a message and persuading their audience, rather than on manipulating forms or simply learning to survive in the target culture. In foreign language learning, a task is defined as “an activity conducted in the foreign language that results in a product with a measurable result such that students can determine for themselves whether or not they have completed the assignment” (Leaver & Kaplan, 2004, p. 47). Nunan (2004) adds to that definition a number of other criteria, emphasizing that, in a task, learners must be focused on expressing meaning, rather than manipulating form. Moreover, like Leaver and Kaplan, Nunan notes that a task should have a “sense of completeness, being able to stand alone as a communicative act in its own right” (p. 4). In task-based foreign language instruction, language becomes a means to an end, rather than an end in and of itself. Debate in the foreign language classroom allows learners to use language as a vehicle for communicating ideas for a meaningful purpose - persuading others, rather than viewing language merely as an object of study (Coyle et al., 2010; Long, 2007; van Lier, 2005).

Debate can be made even more effective when learners have the opportunity to extend their language learning beyond the classroom. To that end, instructors from the U.S. and Russia designed a course based on previously published research demonstrating the effectiveness of debate as a means of promoting Advanced-level proficiency, as well as the emergence of Superior-level functions (Brown, Talalakina, Yakusheva, & Eggett, 2012; Brown, Bown, & Eggett, 2015). An important component of the collaborative course involved the use of teletandems, in which L2 English learners and L2 Russian learners communicated on a weekly basis about topics related to the course to practice their L2 skills.

In this study, we explore the effect of this curriculum design on perceived L2 proficiency gains by examining students’ self-perceptions of language gains in two parallel debate courses: one focusing on L2 English (taught in Russia) and the other on L2 Russian (taught in the United States). Students at both universities followed a parallel curriculum, focused on oral and written debate. Additionally, students were expected to engage in weekly teletandems. A teletandem, as defined by Telles (2015), is a “virtual, collaborative, and autonomous context in which two speakers of different languages use the text, voice, and webcam image resources of VOIP technology (Skype) to help each other learn their native language” (p. 603). Regardless of the platform used, whether Skype or a similar technology, teletandems are traditionally viewed as contributing to the students’ language proficiency (Cardoso & Matos, 2012; Consolo & Furtoso, 2015; Marques Spatti Cavalari, & Aranha, 2016). Similarly, in our project, teletandems helped students practice tasks carefully designed to push their language proficiency and facilitate cultural understanding. To examine the effect of this debate curriculum on students’ perceived language gains, students were asked at the end of the semester to rate their ability to perform a series of functions both at the beginning and end of the course. Presented in the form of Can-Do Statements based on those developed by the National Council of State Supervisors for Languages (NCSSFL) and the American Council on the Teaching of Foreign Languages (ACTFL), these Can-Do Statements permitted insights into how students responded to tasks reflecting different levels of proficiency, i.e., Advanced vs. Superior-level proficiency, and different modes of communication, i.e., interpersonal vs. presentational communication that were practiced in the course. The following research questions guided our study:

RQ1: *Language gains over the course*: Did students feel like they made language gains during a single-semester course focusing on debate?

RQ2a: *Effect of proficiency level and mode of communication on overall performance*: Did students rate their ability to perform at one level (Advanced or Superior functions) or one mode of Can-Do statement (interpersonal vs. presentational) better than another regardless of time? In other words, were some types of Can-Do statements inherently easier for students to perform?

RQ2b: *Effect of proficiency level and mode of communication on language gains*: Did students report more gains on

some types of Can-Do Statements than others, e.g., Can-Do Statements associated with Advanced vs. Superior functions, or with presentational vs. interpersonal statements?

RQ3: *Impact of university*: What differences, if any, were noted in participants' self-assessments between the institutions in the two countries?

This article is organized as follows. First, it offers a theoretical framework for the inclusion of teletandem learning and debate in the foreign language classroom, drawing on recent research on the benefits of such an approach to language teaching and learning. It also introduces the reader to the underlying features of Advanced and Superior proficiency levels as opposed to Novice and Intermediate. Next, it presents the results of the study, examining the effects of a parallel debate course offered at two universities, one in the United States and one in Russia. Before concluding, we present the implications of the study for the classroom in promoting language gains at the Advanced and Superior levels as well as directions for future research on promoting higher levels of language proficiency.

Teletandem in Language Learning

Systematic reviews of online intercultural exchange projects stress the great variety of theoretical paradigms and pedagogical approaches used in such projects, making it difficult to generalize about their effects (Lewis & O'Dowd, 2016). Nevertheless, analyses of the existing evidence suggest that telecollaboration has a positive influence on second language development (Belz, 2003; Chen & Yang, 2014; Vinagre, 2005). However, scholars have expressed concern that the majority of the published research on telecollaborative projects has been carried out between European and North American classrooms (Akiyama & Cunningham, 2018). Thus, the present research aims to fill in the gap in case studies on less common teletandems, in which the differences go beyond 'Western' culture.

An important focus of the research on teletandems concerns questions of culture. For instance, Furstenberg, Levet, English, & Maillat (2001) explore the mentality associated with a foreign culture to show that focusing on culture in the language class can contribute to the development of cross-cultural literacy. Conversely, Ware (2003) concludes that teletandems do not readily promote cultural understanding, emphasizing that a number of factors play into the ultimate success of any telecollaborative project. Belz (2002), for example, in examining telecollaboration through the lens of social realism, suggests that instructors must raise students' awareness of the very concept of intercultural communication and to the varying social norms and practices inherent to particular cultural groups. The importance of promoting cultural literacy notwithstanding, this project focuses primarily on the perceived language gains of the participants. Nevertheless, one of the goals of the teletandem project was to create an environment for cross-cultural understanding as learners debated hot-button social topics.

Proficiency-Based Language Teaching and Debate

The ACTFL Proficiency Guidelines, developed in the 1980s based on the U.S. Civil Service Commission's 1952 Interagency Language Roundtable (ILR) proficiency scale (Herzog, 2003), describe what individuals can do in real-world situations using various language modalities (speaking, writing, listening, and reading). The ACTFL guidelines have become the gold-standard for assessing language skills in the United States and bear some resemblance to the Common European Framework of Reference (CEFR) developed in the late 1990s and early 2000s (Bärenfänger & Tschirner, 2008; see Tschirner, 2012, for correspondences between the ACTFL descriptors and the CEFR levels).

As a framework to assess L2 skills, the Proficiency Guidelines have transformed language teaching by informing pedagogical decisions, from course design to classroom instruction (Omaggio-Hadley & Terry, 2000). As with the CEFR, the ACTFL guidelines have helped shift the focus of language teaching from what learners *know* about the language to what they can actually *do* in the target language, making the ultimate goal of language teaching the development of communicative competence with regards to specific functions at each proficiency level. Consequently, the ideal language classroom of the 21st century now focuses on task-based instruction (Little, 2006). Bygate, Skehan, and Swain (2013) define a task as "an activity which requires learners to use language with emphasis on meaning to obtain an objective" (p. 11). In task-based instruction, learners are

engaged in goal-oriented communication that resembles real-world activities (Ellis, 2003; Pica, 2008; Skehan, 2003). Foreign language educators are increasingly designing tasks that help learners fulfill the global functions at each level of the proficiency scale. Table 1 delineates each of the four major levels of the ACTFL proficiency scale, detailing the global functions of each level, the text type (or organization of discourse), the time frames in which learners can function, their level of accuracy, the range of topics they can discuss (perspective), and their comprehensibility. The characteristics outlined in Table 1 inform both writing and speaking. The separate delineations of each of these characteristics sets the ACTFL guidelines apart from the CEFR in which many of these characteristics are found in the corresponding levels, but without being separately addressed to the same extent, particularly as they pertain to text type.

Table 1
Characteristics of ACTFL Proficiency Levels

	Novice	Intermediate	Advanced	Superior
Text Type:	Lists of words, memorised chunks and phrases	Simple sentences, series of sentences	Paragraph (an organised paragraph with transition words and complex sentences)	Organised series of paragraphs and detailed extended discourse
Timeframe:	Present	Present	All (past, present, future)	All (Past, present, future) Subjunctive/Conditional
Perspective:	Personal	Everything is on the personal level	Can discuss home and local community	National and international themes without taking discussion back onto oneself; idiomatic and may even add in cultural references
Functions:	Can communicate minimally with formulaic and rote utterances, lists and phrases; little functional ability	“Create with language” (simple descriptions and conversations); can pose and answer simple questions; can handle simple transactions	Detailed descriptions; give and explain opinions (without supporting); narrations; give directions; handle a transaction with a complication	Discuss themes in detail and deeply; propose hypotheses and think abstractly; and support opinions
Accuracy	No grammatical control	Emergent control over basic grammar structures in present tense	Sufficient control of grammar across all time frames using grammatical structures which add to the complexity of the language and generic vocabulary with some patterns of errors in the language	No pattern of error in basic structures; specialised vocabulary; can handle linguistic complications
Who can understand them?	May be difficult to understand even by sympathetic interlocutors used to non-native speech	Understood by interlocutors used to dealing with non-native learners of the language	Can be understood by native speakers including those who are not used to non-native speech	Can be understood by native speakers including those who are not used to non-native speech; organisation and accuracy allow listener to focus on message and not on form of language being used

Of relevance for the current discussion are the characteristics that set Advanced and Superior proficiency levels apart from the lower proficiency levels as well as from each other. Indeed, these features provide insight into the reasons why debate and argumentation can help move students into these higher levels of proficiency. The first and most critical feature for our discussion here is function. The Advanced level requires students to be able to discuss topics with substantial detail, providing narrations and descriptions while also offering the beginnings of opinions. The ability to handle details on concrete topics at this level lays the foundation for later shifting to more hypothetical and abstract discussions where opinions must be grounded in evidence, a hallmark of both the Superior proficiency level and good debate technique. Second, in Advanced and Superior level discourse,

THE EFFECTS OF AN EFL AND L2 RUSSIAN TELETANDEM CLASS

language learners shift from talking about topics related to their lives (Novice and Intermediate) to topics relevant to the local community (Advanced) and ultimately to those of national and international importance (Superior). These Superior level topics serve well as the focus of debate, particularly in public discourse. Third, such discussions require the speaker to handle more complex text types, initially in terms of paragraph-level speech (Advanced) before graduating to the multiple paragraphs of extended discourse of Superior-level speech. The final critical feature setting these two higher proficiency levels apart deals with accuracy, where students can still have some patterns of errors but have adequate control of the grammar and vocabulary at the Advanced level to be able to function without miscommunication. Learners are able to implement circumlocution at this level when specific and specialized terminology is lacking. At the Superior level, though some errors may persist, consistent patterns of errors have disappeared. More importantly, Superior-level speakers have developed a specialised vocabulary sufficient to handle in-depth discussions on topics of global relevance.

Supporting and defending opinions is a core task for Superior-level speakers and writers. Thus, the criteria outlined in the ACTFL descriptions of Superior-level L2 users correspond to qualities emphasized in public speaking, debate, and persuasive writing (e.g., writing/speaking in a variety of content areas, cohesive texts of multiple paragraphs, control of a range of grammatical structures, and an extensive vocabulary, allowing the user to select words that reflect subtle differences of meaning). Indeed, progressing towards Superior-level proficiency involves not only improving language skills, but also developing cognitive skills to perform more demanding functions. Indeed, research conducted by Massie (2005) and Connor (1987) identifies the task of argumentation and debate as a valuable strategy for improving both L2 oral and written proficiency, particularly at the Advanced level. Moreover, recent studies by Brown and colleagues (Brown, 2009; Brown, Bown, & Eggett; 2009; Brown & Bown, 2015; Brown, Talalakina, Yakusheva, & Eggett, 2012) have further demonstrated that debate can lead to significant oral proficiency gains within the space of one semester. This burgeoning body of research emphasizes the benefits of carefully designed tasks aligned with proficiency standards.

It is worth noting one further difference highlighted in the research questions, namely the contrast between presentational and interpersonal speaking. This distinction is captured by the CEFR as “spoken production” and “spoken interaction” respectively in the self-assessment grids (“Self-assessment grid – Table 2 (CEFR 3.3)”, n.d.) and in the NTSSC-ACTFL Can-Do Statements. Presentational communication involves one-way communication in which the speaker simply conveys information to a listener. This mode of communication is reflected in many classroom settings where student give formal presentations on a topic to an audience, e.g., classmates. On the other hand, interpersonal speaking involves two-way communication requiring the negotiation of meaning between two or more individuals. Unlike presentational speaking where the speaker maintains control of the discourse by speaking, interpersonal speaking requires both speaking and listening as part of a conversation. Both (or all) individuals must listen to what is being said to be able to respond appropriately to the conversation. While presentational speaking can be prepared in advance, interpersonal speaking by its very nature is more spontaneous in nature making it much more difficult to prepare in advance. One must listen to be able to respond. Moreover, while presentational speaking is often associated with more formal contexts and longer discourse, interpersonal communication is often more informal and typically involves shorter durations of speech in each response and turn taking. While some instructors focus more on conversational exchanges on topics, i.e., interpersonal speaking, others may involve students giving more formal presentations to their classmates.

The ACTFL Proficiency Guidelines, and corresponding assessments have influenced the teaching of languages in the United States for many years, and the effort to translate the guidelines into classroom practice can be traced to the development of the “World Readiness Standards for Learning Languages,” the most recent version of which was published in 2015. As the document states, these serve as a “roadmap to guide learners to develop competence to communicate effectively and interact with cultural understanding.” The Standards outline five goal areas, known as the “5C’s”, that describe the links between communication and culture, as they are applied in making connections and comparisons and in using this competence to participate in local and global communities.

The Standards include:

1. Communication: Communicate effectively in more than one language in order to function in a variety of situations for multiple purposes.
2. Cultures: Interact with cultural competence and understanding.

3. Connections: Connect with other disciplines and acquire information and diverse perspectives in order to use the language to function in academic and career-related situations.
4. Comparisons: Develop insight into the nature of language and culture in order to interact with cultural competence.
5. Communities: Communicate and interact with cultural competence in order to participate in multilingual communities at home and around the world.

A well-designed collaborative debate course can facilitate the achievement of these standards, offering students opportunities to engage in all three modes of communication, as well as to make connections with other disciplines. Incorporating teletandems allows learners to participate in multilingual communities around the world. Not only is such a course poised to facilitate the development of oral proficiency and the achievement of important language learning competencies, but it can also enhance student engagement, motivation, and cultural awareness.

Materials and Methods

The study outlined in this section sought to investigate how the debate course with teletandem interactions using Skype impacted the perceptions of the language gains of the students enrolled in the course at both universities. In this section, we provide an overview of the study. The overview includes a description of the students who participated, the survey used to collect data on the self-perceptions of their progress over the course of the semester, and how the data were analyzed.

Participants

Students participating in the study were undergraduate senior students enrolled in the Global Debate class at their respective institutions. In total, 18 students (6F, 12M) at a large private university in the United States and 19 (13F, 6M) students at a large university in Russia completed the survey. The American students, whose ages ranged from 22 to 25, were all native speakers of English learning Russian as a foreign language (RFL). All were in their fourth year of undergraduate studies majoring in Russian (including those completing a double major in Russian and another field) and most had spent 16 to 22 months abroad in Russian-speaking milieux. The students in Russia, whose ages ranged from 20 to 22, were all native speakers of Russian learning English as a foreign language (EFL) and completing undergraduate degrees in World Economy and International Affairs. Although proficiency was not formally tested before or during the course, both instructors were familiar with the ACTFL proficiency guidelines. It is estimated that students from both groups had proficiency levels ranging from Intermediate High to Superior, with most falling within the Advanced proficiency range, i.e., Advanced Low to Advanced High.

Course Design

The two courses taught in the U.S. and Russia were based on parallel textbooks, *Mastering Russian through Global Debate* (Brown, Balykhina, Talalakina, Bown, & Kurilenko, 2014) and *Mastering English through Global Debate* (Talakina, Brown, Bown, & Eggington, 2014) respectively. The primary objective of these textbooks is to facilitate development of Superior-level language skills via oral debates and written position papers. Although the two textbooks follow a similar structure and contain the same topics (including Economy vs. Environment, Interventionism vs. Isolationism, Wealth Redistribution vs. Self-Reliance, Cultural Preservation vs. Diversity, Security vs. Freedom, and Education vs. Real-World Experience), the exercises and texts within each volume were developed separately. Information about the philosophy and design of the textbooks can be found in Brown and Bown (2015).

Although the teletandem exchanges represented the distinguishing feature of this transnational course, students engaged in a number of other activities, allowing them to meet the “World Readiness Standards for Learning Languages” described earlier. They engaged in interpretive communication as they read and listened to news reports and opinion pieces on the topics of discussion. Learners participated in presentational communication as they presented their arguments for or against particular topics, and in interpersonal communication as they

THE EFFECTS OF AN EFL AND L2 RUSSIAN TELETANDEM CLASS

communicated in their teletandems and discussed opinions and current events in class. Students engaged directly with culture as they discussed perspectives on various current events and the values underlying those perspectives. Connections to disciplines such as economics, history, and political science were important in order to effectively argue for a position. During teletandem exchanges, students made comparisons across cultures as they compared the perspectives underlying such sayings as “from rags to riches” and “из грязи в князи.” Finally, as they participated in teletandem exchanges, learners engaged with multilingual communities beyond their classroom.

Over the course of the semester that this study was conducted, students in Russia and in the United States debated four of the six topics within their respective courses: Environment vs. Ecology, Interventionism vs. Isolationism, Wealth Redistribution vs. Self-Reliance, and Cultural Preservation vs. Diversity. The course syllabi were not identical; that is, the learning outcomes and assignments for each of the two courses varied. For example, the Russian students engaged in more formal individual presentations, while the American students’ presentations were limited to the team debates. What the courses had in common, however, was the use of authentic listening and reading material, discussions and debates, and persuasive writing assignments on each topic. Although the Russian and U.S. students did not engage in team debates via video conferencing, the courses were designed so that students in Russia and in the United States simultaneously discussed the topics listed above. The shared schedule facilitated teletandem exchanges, in which students in the U.S. and Russia participated in focused conversation exchange via the internet.

Once a week, English as a Foreign Language (EFL) students from Russia engaged in a 30-minute conversation exchange with their counterparts learning Russian as a Foreign Language (RFL) in the United States. These one-on-one synchronous exchanges focused on interpersonal communication skills at the Advanced level and beyond. The students were engaged in a once-a-week exchange for 10 weeks with conversation partners rotating each week. Each session lasted 30 minutes with target languages switching after 15 minutes. To better focus student conversations, the instructors identified a series of questions--both in English and in Russian--related to each topic. Sample questions included, “Describe the Russian or U.S. tax system. What kinds of taxes are paid and by whom? Are there any tax exemptions?” Students were instructed to decide in advance on two questions for discussion - one would be posed in English by the native English speakers for the EFL learners to answer, while the Russia-based students would pose a question in Russian to be answered by their U.S. counterparts. Because the topics typically involved research, the pairs of students decided in advance which topics they would discuss. As a follow-up to their conversations and to ensure that students attended to each other’s speech, students then wrote a summary in their L2 of what they had learned from their conversation partner.

Survey: Self-Assessment of language gains during the course

The primary instrument employed in this study was a self-assessment survey designed by two of the authors to examine whether students felt they made progress in their L2, i.e., Russian or English, over the course of the semester. The survey was administered once via Qualtrics at the completion of the course. The first two self-assessment questions directly asked students whether the course helped them with their spoken and written language skills. To respond, students were asked to note their agreement with the statements using a five-point Likert scale where 1= “strongly agree” to 5= “strongly disagree”:

- This course helped me make progress in my spoken language.
- This course helped me make progress in my written language.

Next, students were asked to rate how confident they were in their ability to perform a series of tasks, reflected in Can-Do Statements. They reflected back on what they believed they could do at the beginning and reflecting on where they were at the end of the course. See below for a discussion of the advantages of this type of survey for self-assessment. Their responses were given using a five-point Likert scale:

1. Could not do this even with extensive preparation.
2. Unsure as to whether I could or could not do this.
3. Could do this with extensive preparation.
4. Could do this with minimal preparation.
5. Could do this without any preparation.

The Can-Do statements presented in the survey were derived from the NCSSFL-ACTFL Can-Do Statements (ACTFL, 2013). These Can-Do Statements represent self-assessment checklists that allow learners to assess what they are able “to do” in the L2. According to ACTFL, the “current Can-Do Statements are strategically aligned with the *ACTFL Proficiency Guidelines 2012* and the *ACTFL Performance Descriptors for Language Learners*, both of which remain the standard for assessing foreign language learning in the United States. These Can-Do Statements describe the specific language tasks that learners are likely to perform at various levels of proficiency” (p. 2) in each of the three modes of communication: interpersonal, presentational, and interpretive. An example Can-Do statement from the Superior level is as follows: “I can skillfully relate my point of view to conversations about issues, such as foreign policy, healthcare, or environmental and economic concerns to those made by other speakers.”

The Can-Do statements used in this survey represent presentational and interpersonal tasks at both the Advanced and Superior proficiency levels. A total of 33 separate Can-Do Statements were presented in the survey administered online via Qualtrics, eight Advanced Interpersonal Can-Do statements, eight Superior Interpersonal, six Advanced presentational, and 11 Superior presentational. Samples of each variety are provided in Table 2 with sections highlighted that are relevant for the proficiency level classification:

Table 2
Distribution of Can-Do Statements used in study

	Advanced	Superior
Interpersonal	8 statements Sample: “I could resolve an unexpected complication that arose in a familiar situation.”	8 statements Sample: “I could support my opinions clearly and precisely and construct hypotheses.”
Presentational	6 statements Sample: “I could deliver short presentations on social and cultural topics.”	11 statements Sample: “I could present a viewpoint with supporting arguments on a complex issue.”

As noted, students were asked to respond to their perceived ability to perform each of those Can-Do Statements at the beginning and end of the semester using the five-point Likert scale outlined above. Since students were asked at the end of the course, the wording was changed from “can” to “could” as illustrated in Table 2. The researchers administered the survey just once at the end of the course, employing a “then-now”, i.e., a post + retroflective survey to evaluate students’ perceived gains rather than administering pre- and post-course surveys. In a post + retrospective self-assessment, learners are asked at the conclusion of a program to retroactively evaluate their abilities prior to the learning experience as well as to rate their abilities following the experience (Brown, Dewey, & Cox, 2014). This type of survey design has the advantages of the pretest and posttest model, in which learners are asked to evaluate their abilities in two different sittings—one prior to the learning experience and once again following the learning experience (Meara, 1994). However, research indicates that learners’ perceptions often undergo a significant shift between pretesting and posttesting. The perspective shift usually occurs as a result of learners’ standards changing from pretest to posttest (Gilovich, Kerr, & Medvec, 1993). As Brown, Dewey, and Cox (2014) state, “Prior to the learning experience, [learners] may overestimate or underestimate their abilities due to a lack of experience, but following the experience, they know better what the tasks entail on which they are asked to rate themselves” (p. 265). Moreover, Rohs and Langone (1997) argue that “then-now” assessments allow learners to “[evaluate] themselves with the same standard of measurement or level of understanding on both their posttest responses (how they feel now) and how they felt before the program (then)” (p. 156).

The research literature indicates that self-assessment can promote greater learner awareness and self-regulation. Moreover, involving students in the assessment process can increase learner motivation and participation in the learning process (Dickinson, 1987; Oscarson, 1997; Ross, 1998, 2006). Additionally, self-assessment is relatively easy to design, administer, and score. Self-assessment represents a cost-effective method of evaluating progress, especially when compared with such pricey alternatives as the OPI or Test of Russian as a Foreign Language. Though they do not eliminate the need for certified assessments at different times, self-assessments can help monitor progress between such formal assessments. Finally, as Badstübner and Ecke (2009) note, self-assessment is “representative of students’ perceptions, which in the end, determine a [...] program’s success and survival, perhaps more so than proficiency tests” (p. 42).

Research has also shown that self-assessments are most useful when they are tied to tasks that are familiar to learners or that learners can imagine engaging in (Oscarson, 1997). Moreover, as Oscarson notes, learners are better able to assess their ability in relation to “concrete descriptors of more narrowly defined linguistic situations” (p. 183), such as those used in the ACTFL Can-Do statements employed in this study. This is confirmed by Grahn-Saarinen (2003), who noted that students overestimate their skills when asked to assess their language abilities against abstract standards since they do not understand what type of linguistic knowledge they still lack. However, when asked to rate themselves on their ability to use the language, students have a good understanding of what they can do with the language in different situations and contexts.

Data Analysis

To begin our analysis, subjects’ retroflective (“then”) and post-course (“now”) responses were subjected to a reliability analysis in SPSS since the data are based on self-ratings. Next, to answer the research questions, results of the survey data were analyzed using a series of (Mixed) ANOVAs. Two sets of survey data were examined in this way. First, the assessments students gave for their perceived improvements in both writing and speaking were examined using a one-way ANOVA to determine whether perceptions differed as a function of the university in which each student was enrolled. The second set of data examined were the self-assessments students provided for their ability to perform the 33 Can-Do statements at the beginning (Pre) and end of the course (Post). These data points, namely the Pre and Post, were analyzed using a Mixed ANOVA to determine the effect of the university where students were enrolled, whether students were more likely to improve on one mode of Can-Do statements, e.g., presentational or interpersonal, and whether students were more likely to feel they made improvements on the Can-Do statements targeted for either Advanced or Superior levels as well as any interactions between these factors.

Results

In this section, we begin by presenting a reliability analysis of the students’ self-perceptions of their ability both prior to and following the course before presenting the results of statistical analysis of the self-perception survey data collected from the students at the end of the course and reflecting their perceptions of their ability to perform the various functions both at the outset of the course and at the end of the course.

Reliability of Students’ Responses to Can-Do Survey

Since this study relies on the self-perception of students, the reliability of the Can-Do survey was calculated. Students’ retroflective responses regarding how well they thought they could perform certain tasks at the beginning of the semester and their Post responses (at the end of the semester) were found to be highly reliable and stable (Pre: 37 students; $\alpha=.94$; Post: 37 students; $\alpha=.93$).

Student Self-Perceptions on Progress in Writing and Speaking

A series of one-way ANOVAs were run to determine whether the university where students were enrolled had an effect on the self-reported progress students made for writing and speaking. There was no significant effect for university on the amount of progress students reported in speaking ($F(1,36)=.055, p=.816$). U.S. university students ($M=1.50, SD=.71$) and Russian university students ($M=1.55, SD=.60$) both reported very similar gains in speaking. The means demonstrate that students either agreed (2) or strongly agreed (1) that the course helped them make progress in their spoken language. This is illustrated in Figure 1 below by the blue bars. Here, the closer the bar is to 1, the more students felt the course helped their speaking.

The one-way ANOVA run to determine if university impacted the progress students felt they made in their writing approached but did not reach significance ($F(1,36)=3.955, p=.055$). Although not significant, the Russian students reported making smaller gains in writing ($M=2.70, SD=.98$) as illustrated by the higher bars in Figure 1, than their U.S. university counterparts ($M=2.11, SD=.83$). The lower responses closer to 2 represent agreement with the statement that the course helped them improve their writing, while higher bar responses closer to 3 indicate responses neither agreeing or disagreeing.

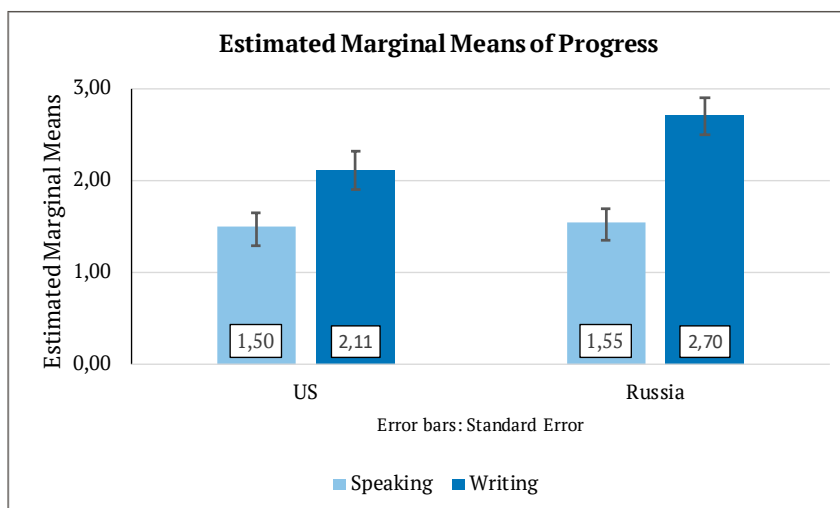


Figure 1. Students' self-report of progress in speaking and writing.

Next, a paired-samples t-test was conducted to compare the perceived progress students made in writing and speaking. Students reported making significantly more progress for speaking ($M=1.53$, $SD=.65$) than for writing ($M=2.42$, $SD=.95$) during the course, $t(37)=-6.39$, $p<.001$, $d=-1.10$, 95% CI $(-1.18, -.61)$. Recall that the closer the rating (and the means of the rating) is to 1, the more students agreed that the course had helped them improve their language skills.

Student Self-Perceptions on Progress on Can-Do Statement Functions

A $2*2*2*2$ mixed-design ANOVA was conducted with time (Pre and Post ratings), Proficiency Level (Advanced and Superior), and Mode (interpersonal and presentational) as within-subjects factors and University (U.S. university and Russian university) as the between-subjects factor. The ANOVA revealed a main effect for all variables. Time was found to be significant, $F(1,35) = 153.66$, $p < .001$, $\eta^2 = .814$, 95% CI $[-1.04, -.75]$, where post-course self-assessments were significantly higher ($M=4.092$, $SD=.35$, 95% CI $[3.975, 4.21]$) than the self-assessments of abilities for the beginning of the course ($M=3.201$, $SD=.499$, 95% CI $[3.035, 3.368]$). This demonstrates that students did perceive that they had made improvements from the beginning to the end of the semester regarding their ability to perform the functions associated with the Can-Do Statements. Proficiency level was also found to be significant, $F(1,35) = 38.32$, $p < .001$, $\eta^2 = .523$, 95% CI $[.242, .479]$, where students rated their ability to perform the Advanced level Can-Do Statements significantly higher ($M=3.83$, $SD=.38$, 95% CI $[3.698, 3.956]$) than their ability to perform the Superior level Can-Do statement functions ($M=3.47$, $SD=.44$, 95% CI $[3.32, 3.61]$). A main effect was also found for the mode of the Can-Do statements $F(1,35) = 5.78$, $p = .022$, $\eta^2 = .142$, 95% CI $[.02, .21]$, where students rated their ability to perform the Interpersonal Can-Do statement functions significantly higher ($M= 3.703$, $SD= .41$, 95% CI $[3.57, 3.84]$) than their ability to perform the presentational level Can-Do statement functions ($M= 3.59$, $SD= .39$, 95% CI $[3.46, 3.72]$) regardless of time, proficiency level, or the university where they were taking the course. University was also found to be significant, $F(1,35) = 5.83$, $p = .021$, $\eta^2 = .143$, 95% CI $[-.545, -.047]$, where Russian students rated their ability on the Can-Do statements higher ($M=3.795$, $SD=.375$, 95% CI $[3.621, 3.968]$) than the U.S. university students ($M= 3.499$, $SD= .373$, 95% CI $[3.32, 3.677]$).

Three additional interactions related to improvement over the course of the semester, i.e., involving time were also found to be significant. First, a two-way interaction between time and proficiency level was significant, $F(1, 35)= 4.396$, $p= .043$, $\eta^2 = .112$ (cf. Figure 2.) Here the label "Beginning" refers to their retroflective estimates of their ability at the beginning of the semester versus how well they rated their ability at the end of the semester.

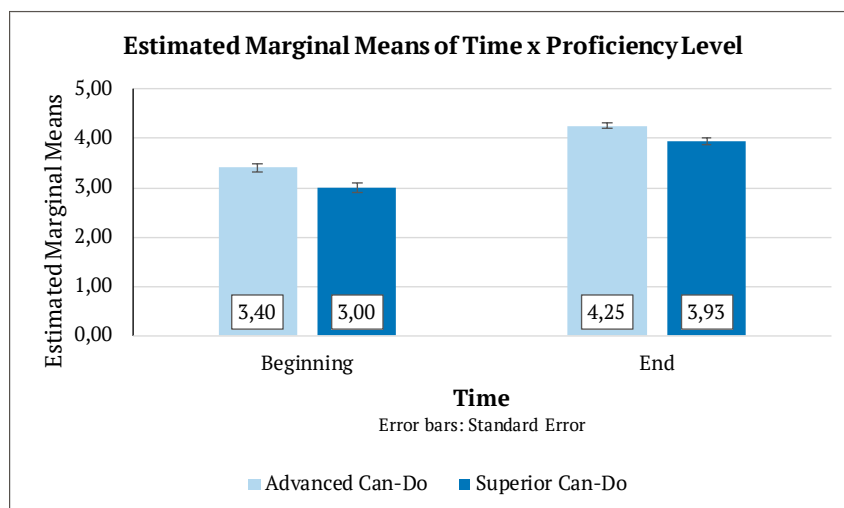


Figure 2. Students' self-report of ability to perform Advanced and Superior Can-Do Statements at the beginning and end of the course.

As illustrated in Figure 2, the Advanced Can-Do statements were rated higher at both the start ($M= 3.4$, $SD= .54$, 94% CI [3.23, 3.58]) and end of the course ($M= 4.251$, $SD= .33$, 95% CI [4.14, 4.36]) than the Superior level Can-Do Statements at the beginning ($M=2.999$, $SD= .55$, 95% CI [4.14, 4.36]) and end ($M= 3.93$, $SD= .44$, 95% CI [3.79, 4.08]). While students noted improvement on both levels of Can-Do Statements, they reported marginally (albeit significantly) more improvement for the Superior Can-Do statements (difference between pre and post means = .934) than for the Advanced statements (difference between pre and post means = .848).

Moreover, a three-way interaction between time x proficiency level x university was also found to be significant, $F(1, 35)= 6.26$, $p=.017$, $\eta^2 = .15$. To further explore the source of this difference, two additional 2*2 within-subject ANOVAs were run for each university to test for an interaction between time and level. No interaction was found between time and proficiency level for the U.S. university students, $F(1, 17) = .84$, $p= .37$. Students reported having made statistically similar gains on both the Advanced (Pre: $M= 3.23$, $SD= .52$; Post: $M= 4.14$, $SD= .37$, mean improvement= .91) and Superior level Can-Do statements (Pre: $M= 2.89$, $SD= .52$; Post: $M= 3.75$, $SD= .47$, mean improvement = .86). However, when the time x level ANOVA was run for the Russian university students, a significant interaction was found, $F(1, 18) = 7.75$, $p= .012$. Students made a slightly, yet nevertheless statistically significant greater improvement on the Superior level Can-Do statements (Pre: $M= 3.11$, $SD= .56$; Post: $M= 4.09$, $SD= .396$, mean improvement = .98) than the Advanced level statements (Pre: $M= 3.59$, $SD= .56$; Post: $M= 4.38$, $SD= .29$, mean improvement = .79). When the improvements between students at the two universities were compared, the Russian students made more improvements on the Superior level statements (although marginally more than the Advanced statements, the difference is significant), while the U.S. university students made marginally more improvements on the Advanced level Can-Do statements (where the difference between the Advanced and Superior-level statements were not significant). Thus, the universities differed in which statements underwent greater improvement.

And finally, a three-way interaction between time x mode (interpersonal vs. presentational) x university was also found to be significant, $F(1, 35)= 4.78$, $p= .036$, $\eta^2 = .12$. To explore the source of the interaction, two additional 2*2 within-subject ANOVAs were run for each university to test for an interaction between time and level. Among the U.S. university students, there was a significant interaction between time and level, $F(1, 17) = 9.92$, $p = .006$, $\eta^2 = .37$. U.S. students reported significantly more improvement on the interpersonal Can-Do statements (Pre: $M= 3.034$, $SD= .55$; Post: $M= 4.00$, $SD= .42$, mean improvement = .97) than the presentational level statements (Pre: $M= 3.04$, $SD= .46$; Post: $M= 3.84$, $SD= .45$, mean improvement = .80) (cf. Figure 4.), although the size of the difference itself is small. It is notable that performance at the start of the semester was similar for the U.S. students on both the presentational and interpersonal statement functions. Students reported improvements on both types of statements, but the improvements for Interpersonal were significantly greater (cf. Figure 4a).

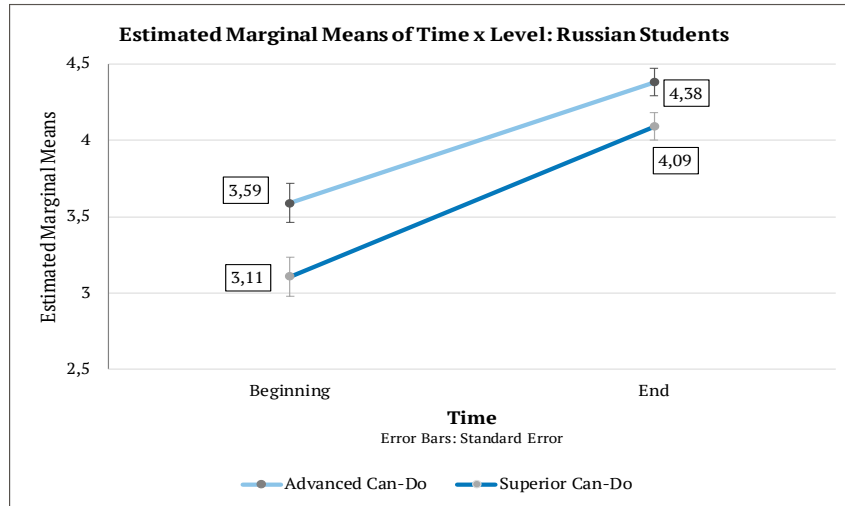


Figure 3a. Russian university students' self-report of ability to perform Advanced and Superior Can-Do Statements at the beginning and end of the course.

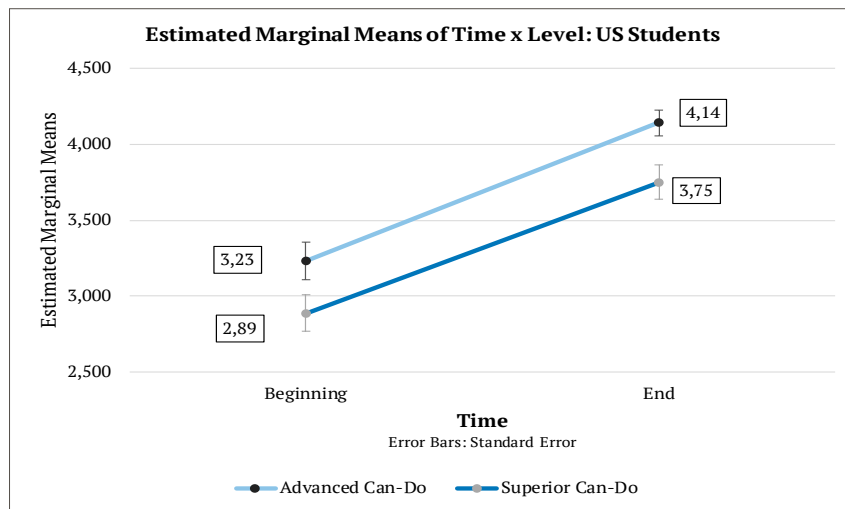


Figure 3b. U.S. university students' self-report of their ability to perform Advanced and Superior Can-Do Statements at the beginning and end of the course.

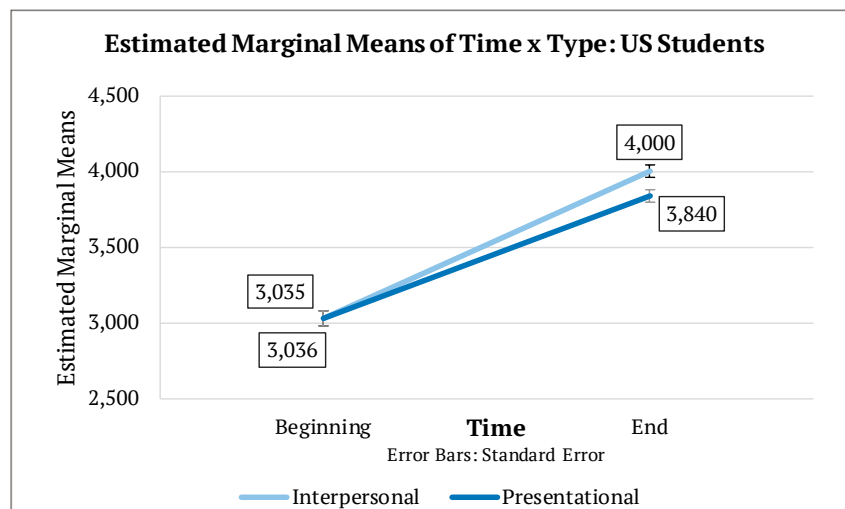


Figure 4a. U.S. university students' self-report on their ability to perform interpersonal and presentational Can-Do Statements at the beginning and end of the course.

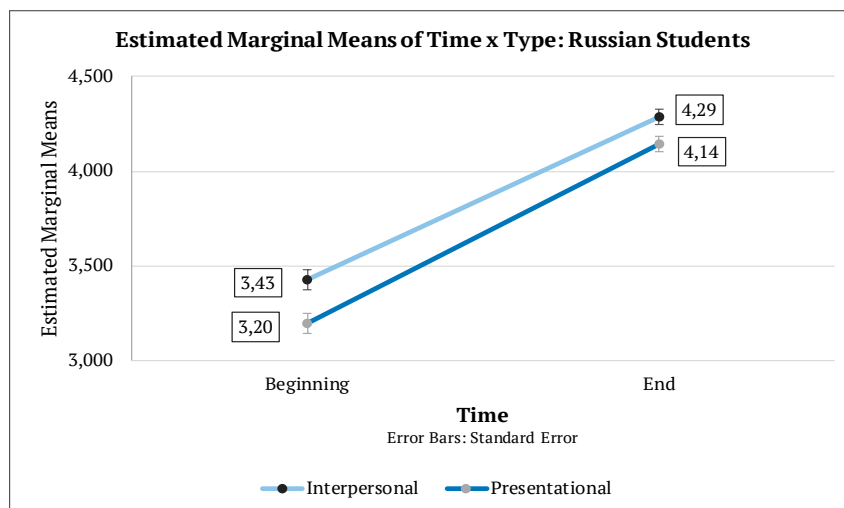


Figure 4b. Russian university students' self-report on their ability to perform interpersonal and presentational Can-Do Statements at the beginning and end of the course.

Among the Russian students, however, no significant difference was found between their self-ratings of the interpersonal and presentational Can-Do statements, $F(1, 18) = .903, p = .36, \eta^2 = .05$. However, it is worth noting that, contrary to the improvement made by the U.S. students (cf. Figure 4b), Russian students reported making marginally more improvement for presentational Can-Do statements (Pre: $M = 3.20, SD = .53$; Post: $M = 4.14, SD = .38$, mean improvement = .95) than for the interpersonal Can-Do statements (Pre: $M = 3.43, SD = .56$; Post: $M = 4.29, SD = .29$, mean improvement = .86).

Discussion

The results discussed above provide insights into student perceptions regarding their proficiency gains over the course of a semester-long interactive debate class. In what follows, we discuss the reliability of the self-ratings provided by the students before systematically responding to each research question.

Reliability of Self-Assessment

Since the data rely on self-reported assessments by students, it is worth discussing the reliability of the data before turning to a discussion of the research questions based on these data. As noted in the results section, the students' pre- and post-assessments were found to be highly stable, providing support for the use of self-assessment data to answer our research questions.

Research question 1: Did students report improving during the course? The most critical question to establish the success of the course is whether it helped students improve their language skills. The results of this study allow us to answer this question in the affirmative. Overall, students reported greater language gains for their speaking than their writing, with the U.S. students assessing their improvement in writing somewhat higher than that of their Russian counterparts. The results are not surprising, as the course focused primarily on the development of oral proficiency through interactive debates and teletandems.

Students likewise reported making improvements with regard to their ability to perform the functions associated with the Can-Do Statements. At the outset of the course, students on average felt they could perform the specified tasks with extensive preparation; but by the end of the course, students on average reported feeling they were now able to perform them with minimal preparation. This marks a substantial improvement in their perception of preparedness to complete the functions specified in the Can-Do statements. Although this study did not measure learners' actual proficiency ratings, these findings do echo Brown's (2009) study in which students made significant gains (as measured by pre- and post-Oral Proficiency Interviews) following a semester-long course focused on debate. This study, taken with other studies of similar courses (Brown, Talalakina, Yakusheva, & Eggett, 2009; Brown, Bown, & Eggett, 2012) suggest that a one-semester course, designed with proficiency

outcomes in mind, can make a difference in students' abilities or perceived abilities without a significant immersion experience. These findings are particularly encouraging in light of the ceiling effect in traditional foreign language programs noted by Rifkin (2005), who asserts that achieving even Advanced-level proficiency may require some kind of intensive immersion experience.

Research question 2: Did students report improving more on one mode and/or level of Can-Do statements?

Research Question 2 asks whether students were better at functions relating to one mode (Interpersonal or Presentational) or level (Advanced or Superior) more than others, and in turn, whether students were more apt to report improvement for one level, i.e., Advanced or Superior level tasks, or mode, i.e., Interpersonal or Presentational. Although the course was focused primarily on the development of Superior-level proficiency, students from both universities generally reported greater facility with and greater gains on tasks related to Advanced-level proficiency. Although the results were statistically significant, in practical terms, the difference is admittedly not substantial. A higher rating for Advanced-level Can-Do statements, if not substantially higher, is not surprising considering that most students, as assessed by the instructors, fell within the Advanced range of proficiency, i.e., between Intermediate-High and Advanced-Mid) at the outset of the course. Since students were enrolled in the course to try to improve their language skills toward the Superior level, it is not surprising that their self-assessment of the skills related to Advanced-level functions would be higher, and that practice in higher-level Superior functions, would result in improvements not only for the Superior-level tasks, but also the Advanced ones.

Theories of L2 learning suggest that output plays an important role in language acquisition (Swain, 1998). Most foreign language educators would recommend that instructors set tasks for learners that will be within their "Zone of Proximal Development" or ZPD (Vygotsky, 1978), defined by Van Patten and Benati (2010) as "the distance between a learner's current ability to use tools to mediate his or her environment and the level of potential development" (p. 152). In other words, tasks should be just beyond the learners' ability to perform them without additional help, provided in the form of scaffolded activities or in interactions with more proficient speakers. This study suggests that learners benefit from being pushed toward the next proficiency level (Thompson, 2008); they consolidate their abilities at their current level of proficiency and begin developing skills at the next level.

With regard to mode of communication, students rated themselves better able to perform interpersonal functions than presentational. In both cases, the responses were midway between "could do this with extensive preparation" and "could do this with minimal preparation," with a marginal yet consistent advantage for the Interpersonal statements as found in the Advanced and Superior comparison. The course's focus on conversational interactions, both in the teletandems and in in-class debates, may contribute to this slight preference.

Based on these two findings, we can answer the initial part of the question in the affirmative, that the targeted proficiency level and communication mode did indeed impact students' perceptions of their ability to perform the functions and tasks outlined in the various Can-Do Statements. Students rated themselves more capable of completing Advanced and Interpersonal tasks. Even if the differences were marginal, they were nevertheless consistent.

Did students make more improvements on the Can-Do Statements based on proficiency level and mode?

To answer this research question, we explored whether self-ratings for proficiency level or communication mode changed over time. While it is true that students consistently rated their ability to perform the Advanced Can-Do Statements higher than their ability to perform the Superior Can-Do functions at both the start and end of the course, students overall reported making more progress on the Superior-level tasks than they made on the Advanced Can-Do Statements (see discussion for RQ3 for clarification on this result). This is not surprising considering the targeted learning outcomes of the course and the focus on debate. The communicative functions of debate, namely discussing topics in depth in order to offer supported opinions and make conjectures about possible consequences, are the very functions that define ACTFL's Superior level. Consequently, the explicit focus on supported opinion, in-depth discussion, and conjecture enabled students to improve their abilities in those functions aligning with the Superior-level Can-Do Statements. Once again, these results accord with previous research on the benefits of language courses focused on debate (Brown, 2009; Brown, Talalakina, Yakusheva, & Eggett, 2009) and underscore the benefits of courses carefully aligned with proficiency outcomes

Research question 3: Effect of University. To this point, the discussion has focused on whether students, regardless of the university at which they were enrolled during the course (and therewith also the language they were learning, either English as an L2 or Russian as an L2 at the Russian and American universities respectively), made improvements. But three questions remain, first as to whether both groups viewed their improvements in speaking and writing similarly; second, whether they perceived their abilities in the various functions outlined by the Can-Do Statements in a similar fashion; and last, whether they benefitted similarly from the debate course.

To answer the first aspect of this question regarding reported gains for speaking and writing, recall that the students at both the U.S. and Russian universities responded similarly to the statement “This course helped me make progress in my spoken language.” Where the two groups of students differed slightly (though not significantly, $p=.055$) was in their response to the corresponding statement for written language improvements. The U.S. students’ responses were, on average, “agree” while the Russian students’ responses were closer to the response “neither agree nor disagree.” Nevertheless, it is positive that students at both universities felt they benefitted from the course, particularly for speaking despite individual differences between the courses at the two universities.

In spite of the fact that the students in Russia consistently rated themselves more favourably than the American students at both the beginning and end of the course,¹ there was no time x university interaction. What this last finding means is that students at both universities made equivalent language gains overall. This is yet again a positive finding since it demonstrates that despite the differences in instructors and in-class approaches, the course design enabled students to note improvements in their oral language skills (the focus of the Can-Do statements). In short, neither group made more gains overall than the other. Thus, despite the difference in locations, instructors, and even languages, students reported that they benefitted similarly from the course design.

That said, the results of this study provide some insights into the unique nature of each group’s progress. Students in Russia reported more improvements for the tasks associated with Superior-level Can-Do statements, while the American students tended to report slightly more improvements on the Advanced-level tasks. This finding may be a natural extension of the higher self-ratings the Russian students gave themselves in comparison to their American counterparts. If the Russian students were already slightly more advanced in their proficiency, they may have been more able to benefit from the practice on the Superior-level tasks addressed in the Can-Do statements. On the other hand, the American students may have found that the practice of the Superior-level functions and tasks contributed to the further development of their proficiency with Advanced-level functions. Again, it must be noted that these differences, while statistically significant, are practically quite small.

A similar difference between students at the two universities was found in the improvement for interpersonal vs. presentational Can-Do statements (time x mode of communication x university). While the American students tended to report greater improvements for the Interpersonal Can-Do statements, the Russian students reported more improvements on the presentational Can-Dos. Again, the differences are very small for practical purposes, but demonstrate slightly different trends in the direction of improvement. These slight differences could be interpreted as a reflection of a different classroom focus. The American students did not engage in formal presentations, beyond the in-class debates, whereas the Russian students engaged in formal presentations in addition to the interactive debates. It should be noted, however, that even though the Russian students reported slightly more improvements in their presentational Can-Do communication, their overall self-assessments showed slightly more confidence in their ability to actually perform the Interpersonal tasks than the presentational ones. The American students, on the other hand, reported similar abilities performing the presentational and interactional Can-Do statements at the outset of the course, and then reported slightly more confidence in their abilities to complete the interactional functions by the end of the course.

¹ It is not clear why this difference at the beginning and end of the study would exist, although a few suggestions can be provided. First, it may be that the students in Russia were indeed more proficient than those in the United States. Since formal proficiency testing, e.g., an Oral Proficiency Interview, was not completed as part of the study, we cannot confirm this. However, based on the instructors’ familiarity with the proficiency levels of both sets of students, this may indeed be the case. This difference may also be based on cultural differences in self-reporting, which have been found elsewhere, particularly comparing Asian and North American students’ self-ratings (e.g., Chen, Lee & Stevenson, 1995). In such studies, American students have been found to be more likely to give higher self-ratings than their fellow students from Asia, underscoring a potential role for a subject’s culture of origin in how they respond to their abilities. The source of the difference is not relevant here since both groups reported feeling that they had made improvements from their relative starting points.

These results suggest that, despite the Russian students' propensity to rate themselves higher in their language skills than the American students, students at both universities benefited from the debate course in terms of language gains, especially gains in spoken language). A closer look at the three-way interactions revealed that the students in each location tended to make slightly different improvements, e.g., the American students improved slightly more on the Advanced and interactional Can-Do Statements, while the Russian students tended to improve slightly more on the Superior and presentational Can-Dos. However, these different trends, while significant, are not necessarily substantial in size and may simply reflect slight differences in classroom culture and focus (for the mode of communication) and the slight difference in self-reported language abilities. What is promising is that such a course allows for overall similar language gains, despite different target languages and administration of the course at the individual institutions. In other words, one group of students did not benefit at the expense of the other.

Implications for the Classroom

Although this study employed self-assessment rather than objective measures of language gains, the results nevertheless indicate that students' confidence in their language abilities can grow significantly following a one-semester language course. As noted above, learners' perceptions of language gain can play a significant role in the success of a particular program (Badstübner & Ecke, 2009). Learners who believe that a course does not lead to improved language ability will be less invested in that course and less likely to recommend it to others. And conversely, if students feel like a class has benefitted their language skills, they will be more likely to recommend the course to others, and those students may in turn feel like their language skills have improved as well.

Moreover, cross-institutional courses involving students learning the L1 of their counterparts at the other university can lead to improvement for both sets of students, even if both sets of students are not at the same proficiency level. In this study, the American students reported slightly lower proficiency in the functions associated with the Can-Do statements than their Russian counterparts. Nevertheless, both sets of students reported increased language ability at the end of the study. The language gains made by the students in the course likely reflect an approach compatible with Krashen's *i+1* hypothesis, which argues that learning occurs along a developmental continuum and, therefore, classroom activities should be just beyond the learner's current stage of development (Krashen, 1988). In this case, the learners had crossed the Advanced-threshold. Making progress required pushing them beyond the functions of the Advanced level into the Superior.

Such an approach, in which students are working just beyond their proficiency level, can also yield results at the students' current level of proficiency. The American group in this study, for instance, reported slightly more progress in the Advanced-level functions than they did in the Superior-level functions. Even though the focus of instruction was not on developing narration and description, performing at the next level, i.e., Superior, helped them to improve in functions at the level below, i.e., Advanced. This suggests the importance of instructional methods and tasks that push learners beyond their comfort level.

Limitations

The primary limitation of this study is its reliance on self-reported data, rather than objective measures of proficiency. Nevertheless, the self-ratings were shown to be reliable and highly stable. Moreover, previous studies by Brown and colleagues (Brown, 2009; Brown, Talalakina, Yakusheva & Eggett) focus on debate in the language classroom, have shown that students make gains on such assessment tools as the Oral Proficiency Interview and the Written Proficiency Test. Future studies might examine the accuracy of students' self-assessments by comparing their responses to more objective measures of gain, such as pre- and post-Oral Proficiency Interviews or Written Proficiency Tests, as used by Brown and colleagues.

Moreover, in relying on self-assessment, this study made use of an unorthodox tool for measuring perceived gains. Rather than administering a survey prior to the course and a second survey after the course, we chose to administer a single "then-now" or post + retroflective survey. A more traditional approach would likely have yielded different results. Nevertheless, evidence suggests that a post + retroflective technique may yield more accurate self-assessments of ability prior to the course (Rohs & Langone, 1997; Lam & Bengo, 2003). In fact, Lam and Bengo (2003), following an extensive review of research on the use of post + retroflective

surveys concluded, “More than three decades of research on post + retrospective method has unequivocally supported this approach over the traditional pretest-posttest approach to measuring change” (p. 78). The design of the study also makes it difficult to identify potential confounding variables. For example, data were lacking on the extent of learners’ participation in course assignments, including the teletandems, as well as on full demographic details on the students’ learning profiles, e.g., length of time studying the L2, time immersed in the L2, etc. Any of these factors may have played a role in their linguistic development and could have provided insights to better interpret the findings.

Future Research

In spite of these limitations, the study does provide insights into the learners’ perceptions of the benefits of a course focused on debate for the development of Advanced and Superior-level functions. Future research can incorporate pre- and post-global proficiency ratings or other tests focused on more specific linguistic skills or knowledge, such as vocabulary or grammar.

Audio or video recordings of student teletandems could prove a rich source of data. Scholars could analyze cross-cultural discourse patterns, negotiation of meaning, error correction, or simply trace development of fluency or vocabulary development over the course of a semester. Qualitative data about students’ experiences both in the course overall as well as in the teletandems could provide further insights into the benefits and challenges of computer-mediated interaction. This study has focused primarily on linguistic gains, but future studies might also focus on issues related to cultural misunderstandings, negotiation, and development of cultural competence. NCSSFL-ACTFL has developed a set of Can-Do statements related to cultural understanding, which would facilitate such a line of research.

The Can-Do statements in this study were introduced at the end of the course. Current research suggests that Can-Do statements can be useful throughout the course as a way of focusing students’ learning and helping them to develop learner autonomy (Lenz, 2004). Future research could examine the effectiveness of introducing Can-Do statements throughout the semester and using them to gauge learning. Scholars can consider how using such self-assessment might improve learners’ accuracy in evaluating their own learning. Whereas the Can-Do statements focus on global tasks and functions, learners likely also make progress in more specific language areas, such as their ability to use transition statements, or build cohesive paragraphs and discourse, or incorporate more specific and specialized vocabulary. Students can be asked to self-assess their progress in these areas, as well as in their ability to perform specific tasks. Moreover, recordings of student presentations and interactions, as well as their written work over the course of the semester can provide a wealth of data, allowing researchers to examine growth in more targeted language features.

Conclusion

The results of this study provide promising insights that curriculum design can indeed impact student proficiency gains over the course of a single semester. As the demand for proficient L2 speakers needed to participate in the global economy continues to rise, there is indeed hope for students to improve their language skills towards Advanced and Superior levels of proficiency. A course drawing on debate and argumentation skills can provide critical help for students to increase their confidence and linguistic preparation to move higher up the proficiency scale. The course outlined here involving both in-class debate preparation and weekly teletandem discussions with native speakers of the students’ L2 is one such way to provide students with the necessary practice on their own home campuses to facilitate the development of professional-level language competence.

References

- Akiyama, Y., & Cunningham, J. D. (2018). Synthesizing the practice of SCMC-based telecollaboration: A scoping review. *Calico Journal*, 35(1), 49-76.
- Badstübner, T., & Ecke, P. (2009). Student expectations, motivations, target language use, and perceived learning

- progress in a summer study abroad program in Germany. *Die Unterrichtspraxis*, 42(1), 41–49.
- Bärenfänger, O., & Tschirner, E. (2008). Language educational policy and the language learning quality management: The Common European Framework of Reference. *Foreign Language Annals*, 41, 81–101. <https://doi.org/10.1111/j.1944-9720.2008.tb03280.x>
- Belz, J. A. (2002). Social dimensions of telecollaborative foreign language study. *Language learning & Technology*, 6(1), 60–81.
- Belz, J. A. (2003). Linguistic perspectives on the development of intercultural competence in telecollaboration. *Language learning & Technology*, 7(2), 68–117.
- Brookhart, S. M., Andolina, M., Zuza, M., & Furman, R. (2004). Minute math: An action research study of student self-assessment. *Educational studies in Mathematics*, 57(2), 213–227.
- 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.
- Brown, N. A., Bown, J., & Eggett, D. L. (2009). Making rapid gains in second language writing: A case study of a third-year Russian language course. *Foreign Language Annals*, 42(3), 424–452.
- Brown, N. A., Dewey, D. P., & Cox, T. L. (2014). Assessing the validity of can-do statements in retrospective (then-now) self-assessment. *Foreign Language Annals*, 47(2), 261–285.
- Brown, N. A., Talalakina, E. V., Yakusheva, I. V., & Eggett, D. L. (2012). Argumentation and debate in the foreign language classroom: Russian and American university students collaborating through new technologies. *Russian Language Journal*, 62, 141–166.
- Brown, T., & Bown, J. (2015). *To advanced proficiency and beyond*. Washington, DC: Georgetown University Press.
- Brown, T., Balykhina, T., Talalakina, E., Bown, J., & Kurilenko, V. (2014). *Mastering Russian through global debate*. Washington, DC: Georgetown University Press.
- Brown, T., Bown, J., & Eggett, D. (2015). Advanced foreign language study through debate. In T. Brown & J. Bown (Eds.), *To advanced proficiency and beyond* (pp. 73–86). Washington, DC: Georgetown University Press.
- Bygate, M., Skehan, P., & Swain, M. (2013). *Researching pedagogic tasks: Second language learning, teaching, and testing*. New York, NY: Routledge.
- Cardoso, T., & Matos, F. (2012). Learning foreign languages in the twenty-first century: An innovating teletandem experiment through Skype. In *Media in Education* (pp. 87–95). New York, NY: Springer.
- Chen, C., Lee, S. & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science*, 6(3), 170–175.
- Chen, J. J., & Yang, S. C. (2014). Fostering foreign language learning through technology-enhanced intercultural projects. *Language Learning & Technology*, 18(1), 57–75.
- Connor, U. (1987). Argumentative patterns in student essays: Cross-cultural differences. In U. Connor & R. Kaplan (Eds.), *Writing across languages: Analysis of L2 text* (pp. 57–71). Reading, MA: AddisonWesley.
- Consolo, D. A., & Furtoso, V. B. (2015). Assessing oral proficiency in computer-assisted foreign language learning: A study in the context of teletandem interactions. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, 31(3), 665–689.
- Coyle, D., Hood, P., & Marsh, D. (2010). *Content and language integrated learning*. Cambridge, UK: Cambridge University Press.
- Darhower, Mark. 2014. Literary discussions and Advanced-Superior speaking functions in the undergraduate language program. *Hispania*, 97(3), 396–412.
- Dickinson, L. (1987). *Self-instruction in language learning*. Cambridge, UK: Cambridge University Press.
- Donato, R. & Brooks, F.B. (2004). Literary discussions and advanced speaking functions: researching the (dis) connection. *Foreign Language Annals*, 37(2), 183–199.
- Ellis, R. (2003). *Task-based language learning and teaching*. New York, NY: Oxford University Press.
- Fernandes, M. & Fontana, D. (1996). Changes in the control beliefs in Portuguese primary school pupils as a consequence of the employment of self-assessment strategies. *British Journal of Educational Psychology*, 66, 301–313.
- Furstenberg, G., Levet, S., English, K., & Maillat, K. (2001). Giving a virtual voice to the silent language of culture: The Cultura project. *Language learning & Technology*, 5(1), 55–102.
- Gilovich, T., Kerr, M., & Medvec, V. H. (1993). Effect of temporal perspective on subjective confidence. *Journal of Personality and Social Psychology*, 64, 552–560. <http://dx.doi.org/10.1037/0022-3514.64.4.552>
- Grahn-Saarinen, A.-M. (2003). Self-assessment, reflection and ‘Can Do’ statements. In A. Hasselgreen (Ed.), *Bergen can-do project* (pp. 57–62). Strasbourg, France: Council of Europe Publishing.
- Herzog, M. (2003). Impact of the proficiency scale and the oral proficiency interview on the foreign language program at the Defense Language Institute Foreign Language Center. *Foreign Language Annals*, 36, 566 –

571. <https://doi.org/10.1111/j.1944-9720.2003.tb02146.x>
- Lam, T., & Bengo, P. (2003). A comparison of three retrospective self-reporting methods of measuring change in instructional practice. *American Journal of Evaluation*, 24, 65–80.
- Leaver, B.L. & Kaplan, M.A. (2004). Task-based instruction in the U.S. Government Slavic language programs. In B. L. Leaver & J. Willis (Eds.), *Task-based instruction in foreign language education: Practices and programs* (pp. 47-66). Washington, D.C.: Georgetown University Press.
- Leaver, B. L., & Shekhtman, B. (Eds.). (2002). *Developing professional-level language proficiency*. Cambridge, UK: Cambridge University Press.
- Lenz, P. (2004). The European language portfolio. In K. Morrow (Ed.) *Insights from the common European framework* (pp. 22-31). Oxford, UK: Oxford University Press.
- Lewis, T., & O'Dowd, R. (2016). Online intercultural exchange and foreign language learning: A systematic review. In *Online Intercultural Exchange* (pp. 35-80). New York, NY: Routledge.
- Little, D. (2006). The Common European Framework of Reference for Languages: Contents, purpose, origin, reception and impact. *Language Teaching*, 39(3), 167-190.
- Long, M. H. (2007). *Problems in SLA*. Mahwah, NJ: Lawrence Erlbaum.
- Marques Spatti Cavalari, S., & Aranha, S. (2016). Teletandem: integrating e-learning into the foreign language classroom. *Acta Scientiarum. Language and Culture*, 38(4), 327-336. <https://doi.org/10.4025/actascilangcult.v38i4.28139>
- Martin, C. (2015). Introduction: Past context, present focus, future directions: Shifting focus from intermediate skills in classroom training to advanced/superior and beyond. In T. Brown & J. Bown (Eds.), *To advanced proficiency and beyond* (pp. xiii-xxiv). Washington, DC: Georgetown University Press.
- Massie, J. (2005). Consideration of context in the CBI course development process. In R. M. Jourdenais & S. E. Springer (Eds.), *Content, tasks and projects in the language classroom: 2004 conference proceedings* (pp. 79–91). Monterey, CA: Monterey Institute of International Studies.
- Meara, P. (1994). The year abroad and its effects. *Language Learning Journal*, 10(1), 32-38.
- Murphy, D., & Evans-Romaine, K. (Eds.). (2016). *Exploring the US language flagship program: Professional competence in a second language by graduation*. Bristol, UK: Multilingual Matters.
- NCSSFL-ACTFL Can-Do Statements. (2015). In *ACTFL*. Retrieved from http://www.actfl.org/sites/default/files/pdfs/Can-Do_Statements_2015.pdf
- Nunan, D. (2004). *Task based language teaching*. Cambridge, UK: Cambridge University Press.
- Omaggio-Hadley, A., & Terry, R. (2000). *Teaching language in context* (3rd ed.). Boston, MA: Heinle & Heinle Publishers.
- Oscarson, M. (1997). Self-assessment of foreign and second language proficiency. *Encyclopedia of Language and Education*, 7, 175-187.
- Pica, T. (2008). Task-based teaching and learning. In B. Spolsky & F. M. Hult (Eds.), *The handbook of educational linguistics* (pp. 523-538). Malden, MA: Blackwell publishing.
- Rohs, F. R., & Langone, C. A. (1997). Increased accuracy in measuring leadership impacts. *Journal of Leadership Studies*, 4(1), 150-158.
- Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, 15, 1–20.
- Ross, S. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research, and Evaluation*, 11, 1–13.
- Skehan, P. (2003). Task-based instruction. *Language Teaching*, 36(1), 1-14.
- Talalakina, E., Brown, T., Bown, J., & Eggington, W. (2014). *Mastering English through global debate*. Washington, DC: Georgetown University Press.
- Telles, J. A. (2015). Learning foreign languages in teletandem: Resources and strategies. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, 31(3), 603-632.
- Thompson, C. (2008). Preparing students for writing and talking about literature. *ADFL Bulletin*, 40(1), 20-23.
- Tschirner, E. (Ed.). (2012). *Aligning frameworks of reference in language testing: The ACTFL Proficiency Guidelines and the Common European Framework of Reference*. Tübingen, Germany: Stauffenburg Verlag.
- van Lier, L. (2005). The Bellman's map: Avoiding the "perfect and absolute blank" in language learning. In R. M. Jourdenais & S. E. Springer (Eds.), *Content, tasks and projects in the language classroom: 2004 Conference proceedings* (pp. 13–21). Monterey, CA: Monterey Institute of International Studies.
- VanPatten, B., & Benati, A.G. (2010). *Key terms in second language acquisition*. London, UK: Continuum International Publishing Group.
- Vinagre, M. (2005). Fostering language learning via email: An English–Spanish exchange. *Computer Assisted*

Language Learning, 18(5), 369-388.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

World-readiness standards for learning languages. (2015). Alexandria, VA: National Standards in Foreign Language Education Project and American Council on the Teaching of Foreign Languages.

Appendix A

Survey Instrument

1. I could use my language to handle a complicated situation.
2. I could participate in conversations on a wide variety of topics that went beyond my everyday life.
3. I could compare and contrast life in different locations and in different times.
4. I could resolve an unexpected complication that arose in a familiar situation.
5. I could conduct or participate in interviews.
6. I could communicate effectively on a wide variety of past, present, and future events.
7. I could exchange general information on topics outside my fields of interest.
8. I could handle a complicated or unexpected turn of events.
9. I could exchange complex information about academic and professional tasks.
10. I could exchange detailed information on topics within and beyond my fields of interest.
11. I could support my opinions clearly and precisely and construct hypotheses.
12. I could discuss complex information in debates or meetings.
13. I could participate with ease in complex discussions with multiple participants on a wide variety of topics.
14. I could use my language persuasively to advocate a point of view that was not necessarily my own.
15. I could tailor language to a variety of audiences by adapting my speech and register in culturally authentic ways.
16. I could communicate skillfully and succinctly, often using cultural and historical references to say less and mean more.
17. I could make a presentation on events, activities, and topics of particular interest.
18. I could present my point of view and provide reasons to support it.
19. I could deliver short presentations on a number of academic and workplace topics.
20. I could deliver short presentations on social and cultural topics.
21. I could explain issues of public and community interest, including different viewpoints.
22. I could deliver presentations for a specific audience.
23. I could present information about events of public or personal interest.
24. I could convey my ideas and elaborate on a variety of academic topics.
25. I could give presentations with ease and detail on a wide variety of topics related to professional interests.
26. I could present complex information on many concrete topics and related issues.
27. I could present a viewpoint with supporting arguments on a complex issue.
28. I could use appropriate presentational conventions and strategies.
29. I could give a clearly articulated and well-structured presentation on a complex topic or issue.
30. I could adapt the language in my presentation for casual, professional, or general audiences.
31. I could depart from the prepared text of my presentation when appropriate.
32. I could present skillfully and with accuracy, efficiency, and effectiveness in a variety of settings.
33. I could tailor my presentation to engage an audience whose attitudes and culture may differ from my own.