

Research Approaches to Second Language Acquisition

*Proceedings of the 2018 Second Language Acquisition
Graduate Student Symposium*

Edited by Antonio Alejandro Perez Belda, Hadley Galbraith, Kevin Josephs, Angela Pico Pinto, Evelyn Pulkowski, Kezia Walker-Cecil, and Caolimeng Wuxiha



CARLA | Center for Advanced
Research on Language
Acquisition

UNIVERSITY OF MINNESOTA

Research Approaches to Second Language Acquisition

Proceedings of the 2018 Second Language Acquisition Graduate Student Symposium

Edited by

Antonio Alejandro Perez Belda

Hadley Galbraith

Kevin Josephs

Angela Pico Pinto

Evelyn Pulkowski

Kezia Walker-Cecil

Caolimeng Wuxiha

Research Approaches to Second Language Acquisition: Proceedings of the 2018 Second Language Acquisition Graduate Student Symposium

© 2019 Regents of the University of Minnesota.
All rights reserved.

Produced by
Center for Advanced Research on Language Acquisition
University of Minnesota
140 University International Center
331 17th Avenue Southeast
Minneapolis, MN 55414
USA
612-626-8600
carla@umn.edu
<https://carla.umn.edu/>

Desktop Publishing and Cover Design: Elizabeth Hellebuyck

Cover Photo: Copyright © 2004 Regents of the University of Minnesota

The University of Minnesota is an equal opportunity educator and employer.

This publication is available in alternative formats upon request. Direct requests to the CARLA office at carla@umn.edu.

Acknowledgements

We would like to thank the Center of Advanced Research on Language Acquisition (CARLA) for publishing the proceedings. We are especially grateful for the help of Dr. Kate Paesani, director of CARLA, who provided the resources for the editing of the proceedings and offered detailed and professional guidance throughout the editing process. Without her help, it would have been very challenging for us to finish the project successfully.

We are grateful for the help of Dr. Judith Liskin-Gasparro, associate professor emerita in the Department of Spanish and Portuguese at the University of Iowa, for establishing the specific timeline of the editing process. We thank her for helping us get the proceedings off the ground.

We are grateful for the time and effort of the anonymous reviewers who reviewed the submitted manuscripts. The academic work presented at the 2018 Second Language Acquisition Graduate Student Symposium made the proceedings possible. We would like to thank the parties who financially supported the symposium: the Graduate College of the University of Iowa, the Division of World Languages, Literatures and Cultures, and the Confucius Institute of the University of Iowa.

Table of Contents

Peer Reviewing in L2 Spanish Classrooms: Action Research <i>Emilia Illana-Mahiques, University of Iowa</i>	1
Investigating an L2 French Writer’s Performance on an Integrated Task <i>Anna Evans, University of Iowa</i>	25
A VOT Study on the Acquisition of English Stop Contrasts by Marathi Speakers: The Mutual Influence of L1 and L2 <i>Ari Natarina, University of Iowa</i>	39
An Exploration of L2 Chinese learners’ Perceptions of Interacting with Recap <i>Runqing Qi, University of Iowa</i>	65
“Her Russian Broke my Brain”: The Linguistic Influence of English on the Vocabulary of a Russian Heritage Speaker <i>Janet Stewart, University of Minnesota</i>	85
A Mixed Methods Evaluation of an Urban K–12 World Languages Program <i>Reuben Vyn, University of Iowa</i>	103
About the Authors	129

Peer Reviewing in L2 Spanish Classrooms: Action Research

Emilia Illana-Mahiques, University of Iowa

Peer feedback is a widely used tool across different contexts and levels, including in second language (L2) learning. Despite being a well-accepted practice in L2 educational settings, much is unknown about best peer review practices and how training can maximize learners' critical and analytical skills. Through the lens of action research, we share in this study learning experiences about conducting online peer review and the emerging understanding of how to implement effective training sessions in Intermediate Spanish II classes. The instructional practices were proposed as part of an L2 writing project on personal narratives. The project was developed and refined throughout three consecutive semesters in Spring, Summer, and Fall of 2017. The specific changes implemented from one semester to another, aiming to help students improve their peer reviewing skills, are elements of interest to this study. The training activities that worked best in preparing students to become effective reviewers are described in detail. Ultimately, the purpose of this study is to contribute to a better understanding of what constitutes effective practices in peer review training.

Second language (L2) educators are faced with the challenges of how to best implement writing in the classroom and what procedures maximize student writing ability. Traditionally, L2 instructors have focused on assessing learners' writing as the final outcome of L2 learning, paying little attention to alternative practices that foster processes of learning to write and limiting the responses to writing to the teacher's role only.

More recently, L2 researchers and educators have considered a variety of new practices to maximize writing ability. Some common activities that emphasize L2 learning include peer response groups, collaborative learning, and other small group work activities (Hyland & Hyland, 2006b; Rollinson, 2005; Storch, 2005). Among these, peer review arises as an important part of process-oriented writing instruction (Paulus, 1999).

Peer review refers to students working in dyads or small groups to critique and give feedback on one another's work. The term peer review has been used interchangeably with peer response and peer feedback to denote the process of providing peer assistance for improving writing skills in the form of written feedback, oral feedback, or both written and oral feedback (Gielen, Peeters, Dochy, Onghena, & Struyven, 2010). The positive effects of implementing peer

feedback in the classroom have been widely accepted in research literature and in many composition classrooms (Hyland & Hyland, 2006a). The paragraphs below explore some of the relevant literature.

Literature Review

Findings on the Effectiveness of Peer Review

Peer feedback is seen as a tool that facilitates learning by which students gain autonomy and control over their work. Instead of a passive dependence on a teacher's feedback, they gain a greater voice by engaging in the process of responding to each other's work. Moreover, students also develop the ability to analyze and critically evaluate their own work as writers (Couzijn & Rijlaarsdam, 2005) as well as revise their text based on the peer's feedback (Hyland, 2000; Mendonça & Johnson, 1994).

Engaging in peer review practices helps students develop a sense of audience. Learners take into consideration the characteristics and demands of their audience and learn to process their own work from the perspective of a reader (Cho & MacArthur, 2010). In light of the reader's reactions and responses, writers are empowered not only to better conceptualize their own ideas and make necessary adjustments (Mendonça & Johnson, 1994), but also to expand their perspectives and explore ideas different from their own (Cho & Cho, 2011; Paulus, 1999).

Peer review also facilitates meaningful interactions among reviewers. Students learn to develop more effective ways of analyzing and revising writing by working together in pairs or small groups (Cho & MacArthur, 2010; Zhao, 2010). In general, feedback and revision are strongly recommended tools for leading students to adopt an active role in learning to write (Mendonça & Johnson, 1994; Nystrand, 1986). Discussions in pairs or small groups enhance interaction, negotiation, and the exchange of meaning. Students are able to scaffold each other and build new knowledge that contributes to the process of learning to write in the L2 (de Guerrero & Villamil, 2000; Nystrand, 1986; Wigglesworth & Storch, 2012).

Ultimately, the methodology is well-established in instruction because it provides learners with a range of benefits (Liu & Hansen, 2002) that lead students to improve their written performance (de Guerrero & Villamil, 2000). The benefits are also evident from a logistical perspective: Use of peer review reduces teachers' workload (Dunlap, 2005), facilitates

2 • Research Approaches to Second Language Acquisition

better class dynamics by promoting collaboration between learners with different proficiency levels (Allen & Mills, 2016; Mendonça & Johnson, 1994), and engages learners in all three modes of communication (interpretive, interpersonal, and presentational) within the same task. As Levi Altstaedter (2016) affirmed, learners “write in the presentational mode,” then “read their peers’ drafts in the interpretive mode,” and finally “provide feedback in the interpersonal mode” (p. 2).

Concerns about Peer Feedback and Potential Disadvantages

Against the enthusiastic claims supporting the benefits of peer response, other researchers report negative perspectives from their participants (Leki, 1990; Nelson & Murphy, 1992, 1993; Zhang, 1995). Some of the main concerns emerge in multicultural groups, where different backgrounds along with cultural values impact the dynamics of the peer review. Specifically, students from collectivist cultures (e.g., Chinese, Japanese) may experience cultural inhibitions, which may lead them to avoid disagreement and criticism on peers’ text (Carson & Nelson, 1996).

Beliefs, preconceptions, or other affective factors have also been shown to have a direct effect on learners’ attitudes and behavior during peer review sessions. Learners with negative perceptions towards peer review may act too harshly with their partners or make comments that are too critical (Nelson & Murphy, 1992). In turn, students may not easily accept the idea that their peers, who have the same status and are also learning the L2, are sufficiently qualified to critique their writing (Nelson & Murphy, 1993). This leads to the tendency of placing more trust in teachers’ comments and feeling reserved about accepting suggestions from peers and acknowledging the quality of their feedback (Connor & Asenavage, 1994; Nelson & Murphy, 1993; Zhang, 1995).

Students’ inconsistent quality of peer feedback also has a direct impact on the efficiency of the peer review sessions (Leki, 1990). Issues regarding learners’ limited proficiency in the L2 (Leki, 1990), the quality of peer-to-peer interactions (Mendonça & Johnson, 1994; Papadopoulos, Lagkas, & Demetriadis, 2012), and the type of feedback students use in their comments (Gielen et al. 2010; Strijbos, Narciss, & Dünnebier, 2010) often prevent students from giving specific and useful feedback. For example, reviewers often misunderstand the

writer's intentions, provide vague feedback, or simply give formulaic feedback when they do not know what else to say (Leki, 1990). Other commenters show a tendency to provide feedback on surface level errors only (i.e., grammar errors or mechanics), rather than proposing a wider variety of comments that include more global and meaning-level aspects (Connor & Asenavage, 1994; Nelson & Murphy, 1992, 1993; Tsui & Ng, 2000). As suggested, either students do not have enough confidence to propose major changes, do not have the cognitive skills required to identify macro-level flaws (Tsui & Ng, 2000), or lack the necessary language proficiency to make insightful comments (Leki, 1990). Overall, researchers in the field are called to recognize the challenging factors associated with peer review and the impact of these factors on the process. Recognizing the challenges and limitations of peer review is a crucial step for identifying potential solutions and practices that may yield more efficient pedagogical practices. In particular, a solution widely accepted to counterbalance the negative effects of peer review is providing students with training (Berg, 1999b; Hu, 2005; Min, 2005).

The Role of Training

The limitations discussed above may jeopardize the effectiveness of introducing peer review in L2 classrooms. However, authors who discuss such limitations do not always mention whether or not any training preceded the peer review sessions. They also do not address what specific procedures were applied to guide learners through the stages of reviewing and revising. To date, researchers have recognized the crucial role of training and agree that training is indispensable for the successful development of peer review activities in the classroom. Specifically, researchers such as Zhu (1995), Berg (1999a, 1999b), and Min (2005, 2006) argue that peer feedback is only beneficial when students receive guidance on how to provide feedback to each other.

The positive effects of peer review training have been extensively demonstrated in the literature. As a result, researchers agree on the effectiveness of peer review with training as addressing the various concerns raised above, including the behavioral challenges of multicultural groups (Vorobel & Kim, 2014); learners' cultural inhibitions within communitarian contexts (e.g., Asian countries) (Hu, 2005; Kamimura, 2006; Min, 2005); students' negative attitudes towards peer review activities (Stanley, 1992; Zhu, 1995); the rejection of peer suggestions in favor of instructor feedback (Berg, 1999b; Min, 2006); and the

poor quality of comments given on peer text by students (Lam 2010; Levi Altstaedter, 2016; Min, 2005, 2008; Stanley, 1992).

Approaches Using Peer Review Training

Despite agreement on the positive influence of training, there is little consistency on how to better implement peer review practices. Different researchers have proposed a wide variety of methods and strategies for training students. Briefly mentioned, some of the main approaches include: evaluating written samples and discussing reviewing and revising strategies (Stanley, 1992); analyzing the impact of peer review comments in a final draft (Berg, 1999a); holding one-on-one instructor-student conferences (Min, 2005, 2006; Zhu, 1995); and modelling peer review in the classroom (Berg, 1999a; Hu, 2005; Min, 2006).

Researchers do not always describe the procedures involved in the methodology of training approaches. Berg (1999a) and Hu (2005) are the most comprehensive studies that detail how instructors can implement effective peer review training. Berg (1999a) provided a set of guidelines to prepare English as a Second Language (ESL) students for peer response. The author presented and expanded on a total of 11 guidelines.

Whereas Berg's study (1999a) included helpful suggestions and examples, the guidelines remain somewhat general in terms of designing specific activities that apply to the L2 learning context (not ESL). Hu (2005) addressed this issue in a three-year action research study, providing instructors with a list of activities to train students for face-to-face peer review sessions. Hu's study (2005), however, took place in an intensive ESL program in China, which aimed to prepare students for tertiary education. The present study differs from that of Berg (1999a) and Hu (2005) in that it aims to train students for online peer review sessions (not face-to-face) in the L2 Spanish learning context (not ESL), and in an American public university. This action research study is an attempt to confirm the usefulness of the activities suggested in previous studies (Hu, 2005; Berg, 1999a), incorporate new ones as needed, and report on the most significant experiences from one semester to the next. Ultimately, the goal of this study is twofold: first, to expand the repertoire of training activities that instructors may use and adapt in their own classes; and second, to share learning about how to optimize the pedagogical benefits

of online peer review sessions, particularly at the U.S. university level in an L2 Spanish learning context.

Methodology

Research Method

A reflective process coupled with instructional practices repeated from one semester to another for three semesters set the foundation for this exploratory action research project. Action research is grounded in a qualitative research paradigm and it involves the collection of data that is relevant for professional development (Nasrollahi, Krish, & Noor, 2012). The research methodology arises as a valuable tool for instructors to improve their teaching skills as well as gain greater awareness about their teaching practices, their classrooms, and their students. As stated in Nasrollahi et al. (2012), action research “is also about how teacher [sic] can do things better and how teacher [sic] can change their instruction to impact students” (p. 1875). Thus, the purpose of action research is to shed light on effective peer review practices while advancing professional development in L2 teaching practices (Carr & Kemmis, 2003).

Population and Context of the Study

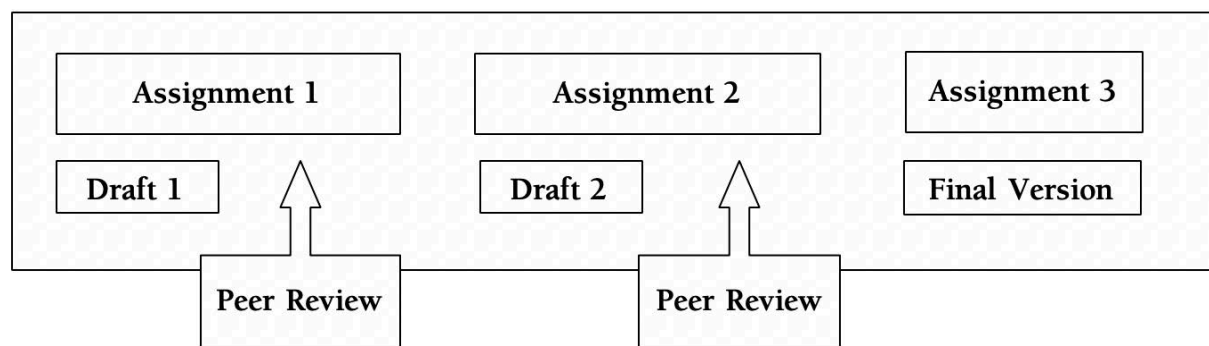
The data for this study were gathered at a large public university in the Midwest. Various implementations of the peer review project occurred in three different sections of an Intermediate Spanish II course. This lower-division course is intended to reinforce grammar and vocabulary, as well as strengthen students’ writing and speaking abilities through a variety of interactive tasks, assignments, and projects. The overarching goal is to develop proficiency and promote acquisition of Spanish language skills (i.e., reading, writing, listening, and speaking) in a variety of communicative contexts.

The pedagogy adopted in Intermediate Spanish II courses at this university may be described as communicative, task-based, and theme-structured. The course curriculum’s focus is on reading culture-related passages, writing a short summary, closely analyzing and discussing the content of the text, and peer reviewing each other’s summaries before moving on to other conversational tasks. These activities occur on a weekly basis and emphasize reading and writing as a social act. They are an excellent preparation for longer L2 writing assignments or future writing courses in Spanish.

Based on the requirements and expectations of the curriculum, a month-duration final L2 writing project was included in the course. Students were asked to write a personal narrative that was engaging and that brought to life an interesting story about themselves. Topics could be a memorable event or an experience that deeply affected their lives or their motivations in life. The project had to be a strong mix of story (narrative), emotion, and reflection. In completing the project, students had to write the stories individually and then share them with their classmates to get feedback for improvement.

In each of the implementations (Spring, Summer, and Fall 2017), the project consisted of a total of three assignments completed over the span of one month: Draft 1, Draft 2, and Final Version. As the assignments alternated with peer review sessions, the procedures followed throughout each assignment were the same: writing a draft, peer reviewing, revising the draft, and submitting the edited draft as fulfillment of an assignment. The structure of the overall project may be better explained in the image below.

Figure 1. Structure of the Month-Duration L2 Writing Project



General Research Procedures: Data Collection and Analysis

The project took place in pre-selected Intermediate Spanish II classes in the Spring semester of 2017 and continued until the Fall semester of 2017, involving three different groups of L2 Spanish learners. The organization of the project can be grouped in two main areas: L2 writing project design and peer review training. First, the L2 writing project design was carried out by the researcher prior the implementation of the project. The steps included: (1) identifying peer review as a pedagogical practice; (2) gathering further information about peer review by reading relevant literature; (3) asking professors about their experiences and conducting observations of peer review activities in upper-level writing courses; (4) obtaining information about online peer

review software and platforms; (5) developing an L2 writing project and adapting the materials to the online peer review setting (i.e., Turnitin Feedback Studio); and (6) making logistic decisions about the timing and deadlines of the project. Overall, this procedure involved making many decisions about the structure and format of the project; the focus, content, and organizations of the tasks; as well as other logistic issues related to length, deadlines, and rules for the project. The first part of the results section expands on some of the most relevant decisions that impacted the three peer review implementations.

The researcher, together with the class instructor, headed student peer review training. After a reflective process of collecting and analyzing data obtained in each project implementation, the specific activities of the L2 writing project were changed and revised as necessary. The reflective process involved reiterative steps of collecting and analyzing data from students' participation in the peer review sessions to track progress of project implementation. The steps included: (1) acknowledging that the peer review training needed improvement; (2) collecting data during implementation of the project (i.e., reflective journals, students' written feedback in Draft 1 and Draft 2 of the project, and student opinions collected in informal conferences or in end-of-the-semester questionnaires); (3) analyzing the data and sharing the results with another researcher experienced in the field; (4) making changes and revisions to the project based on developing understanding about peer review and other relevant information from the literature; and (5) repeating steps 1 to 4 in the following two semesters (Summer 2017 and Fall 2017).

Overall, the five reflective steps presented above include iterative cycles of reflecting, planning, acting, and observing. As these cycles worked in sequence, they offered insights into how students learn and how instructors may improve their instructional practices (Carr & Kemmis, 2003). These insights and experiences are particularly relevant in action research. As Nasrollahi et al. (2012) affirm, "language teachers who engage in action research are developing their professional judgment and autonomy [which] ... implies a different way of generating knowledge" (p. 1875).

To briefly expand on the process of data collection (step 2), the researcher noted several observations for each student through the use of journaling. Such topics included students' behavior throughout the project, the type and quality of comments provided by students during

peer review, as well as potential problems and solutions for future implementations. Students' written feedback in the two drafts offered informative data about their peer reviewing skills. There were also informal conferences with students through office hour appointments that had been set up to talk about each student's overall performance in the course. These conversations included questions related to how they felt about the project, how helpful they found the peer reviewing, and whether they thought the project was any different from the short face-to-face peer reviewing activities that were practiced in class. The end-of-the-year questionnaire provided insight into how the students thought the project could be improved upon by asking students to anonymously write what they liked the most about the project, what they liked the least, and what they would want done differently.

Because of space limitations and the primary interest of action research methodology on the teachers' experiences, practices, and pedagogical insights, this study does not provide further information on the data collection and analyses conducted. As shown in the action research project developed in Hu (2005), it is more interesting in action research to highlight the conclusions and describe the specific pedagogical practices that result from observing learners' behavior, modifying teaching practices, and reflecting on them. With this consideration in mind, the following subsections explain the processes, the reflections and conclusions drawn from the data, and the actions taken to solve issues related to peer review.

Research Questions

The research questions that guided this action research involved ongoing self-reflection and were not meant to have a specific answer. Instead, they were to propel better pedagogical practices.

The research questions were:

1. How can peer review training practices be improved so that they positively affect learners' behaviors, the comments they make, and their overall engagement in the project?
2. How do students react to and what behavior and opinions do they hold about the project after improving the peer review training sessions from the previous semester?

Results

The results of this study expand on the relevant conclusions drawn from the data, and the decisions and procedures taken to successfully enrich the L2 writing project. As mentioned in

the previous sections, two areas are considered: the L2 writing project design and the peer review training.

L2 Writing Project Design

Designing the L2 writing project involved making many decisions, such as how much time to allot for the project, how much time to dedicate to each of the three assignments, and which peer review platform to use. It was decided that there be one month to complete the project, to dedicate at least one week to each of the three assignments during that month, and to use the online peer review platform Turnitin Feedback Studio. The platform allows students not only to give general feedback, but also to highlight and give detailed in-text comments. To facilitate the peer review task, students were allowed to use English, Spanish, or both to make their comments. This flexibility encouraged students to write longer and more thoughtful comments, especially when students encountered difficulties expressing their ideas in the foreign language (Ho & Savignon, 2007; Yang et al., 2006).

For each peer review session, students were randomly assigned so that no two students gave and received feedback from one another. For example, when a student was assigned a draft to peer review, student A was assigned to review the paper of student B, whereas student B was asked to review the paper of student C. This process ($A \rightarrow B \rightarrow C$) allows for anonymity—an important factor that encourages students to comment more and provide more honest feedback and eliminates the situation of students accepting or rejecting comments based on superficial perceptions of the reviewer (Cho & Cho, 2011; Strijbos et al., 2010). Anonymity and random assignment were arranged manually according to students' L2 proficiency level using Turnitin Feedback Studio. This helped boost students' perceived relative proficiency, as well as their confidence in their own ability to provide feedback to their peers (Allen & Mills, 2016).

In each of the three implementations of the project, students had to follow the questions included in the peer review guidelines. To counteract their tendency to focus exclusively on surface-level issues (Hu, 2005; Nelson & Murphy, 1992; Paulus, 1999), the list of questions was divided into two sections. The first section included global or macro-level issues (e.g., content, organization, development, audience, and voice), which were considered of higher priority. The

second section included surface or micro-level issues (e.g., language use, grammar, vocabulary choice, connectors), which were considered lower priority.

The guidelines were significantly changed over the semesters to encourage higher-quality comments. All questions in the guidelines were revised to make them as clear and transparent as possible. First, we made sure to avoid ambiguous and less insightful questions (Nilson, 2003). Then, because the questions were getting too long and elaborated, we added a summarized version of the questions by stating them in terms of goals. As shown in the excerpt from Table 1, both formats were displayed in the guidelines. During informal conferences, students expressed that they were able to work comfortably with this format. The goals allowed for a quick and straightforward understanding of the criteria, and the questions were mainly used whenever they felt they needed more information.

Table 1. Extract of Goals and Questions from the Peer Review Guidelines

Goals	Questions
Title: Captures reader’s curiosity. Related to the story.	Title: To what extent does the title capture your interest or curiosity? How do you think the writer can improve the title?
Beginning of the text: Includes background information. It should engage and captivate the reader.	Beginning of the text: To what extent is there sufficient background information? Does the narrative start off in a way that is engaging and makes you want to keep reading?

Although the aforementioned changes established the conditions for more optimal peer review sessions, the sessions needed to be coupled with appropriate training. As suggested in the literature, developing learners’ critical and analytical skills is necessary to promote successful peer reviews (Berg, 1999a; Hu, 2005; Min 2005). The second area of focus in this exploratory action research is to share our emerging understanding about peer review training and how the activities, strategies, and procedures followed helped the students work toward becoming more autonomous and successful reviewers.

Peer Review Training

Different training activities were implemented throughout the three semesters to guide students toward performing more effective peer review sessions. The conclusions drawn from analysis and reflection of the data are presented below. Table 2 briefly summarizes the information about the activities developed each semester. S1, S2, and S3 refer to each of the semesters (Semester 1,

Semester 2, and Semester 3). The full range of activities is listed, and the checkmark (✓) indicates the semester(s) in which the activities were carried out.

Table 2. Training Activities

	S1	S2	S3
Presentation of the project			
• Explaining the assignment: what students are expected to do, the different stages of the assignment, and time allotted for each stage	✓	✓	✓
• Providing focused instruction that is purposefully aligned with the dimensions of the guidelines		✓	✓
• Contextualizing and building the personal narrative from pictures that students had taken			✓
Awareness-raising			
• Justifying the role of the peer review activity	✓	✓	✓
• Summarizing the research findings on peer review	✓	✓	✓
• Examining examples of peer review as a common technique among colleagues, professionals, in other common-life situations		✓	✓
• Discussing the general benefits and challenges of peer review		✓	✓
• Discussing the specific challenges of peer review in the L2 context and proposing alternatives and solutions		✓	✓
• Establishing a set of principles and values of peer review for the class			✓
Guidelines			
• Providing the peer review guidelines	✓	✓	✓
• Discussing the important role of the peer review guidelines in the reviewing stage		✓	✓
• Analyzing questions in terms of their format and their foci of attention		✓	✓
• Reading the guidelines in class		✓	✓
• Modeling use of the peer review guidelines		✓	
• Practicing the guidelines through hands-on activities			✓
Feedback comments			
• Explaining the typology of feedback comments	✓	✓	✓
• Modeling giving feedback on a sample essay		✓	
• Classifying a set of feedback comments by their typology			✓
• Practicing commenting based on hypothetical situations			✓
• Practicing with adapted sample essays and excerpts			✓
Technology training			
• Providing a handout with step-by-step instructions of how the software works	✓	✓	✓
• Briefly demonstrating how to access the peer review software		✓	✓
• Going to a computer room and, in small groups, giving online feedback to a sample essay			✓

Analysis and Reflections on the First Implementation

The first implementation of the L2 writing project in Fall 2016 was the least successful. The final drafts did not meet expectations, and the students' comments to their peers were far from helpful. The comments were limited as a result of being too short, ambiguous, and not insightful. Students also avoided critiquing their peers' paper. Instead, they gave positive formulaic feedback that lacked specificity and justification such as "well done" or "good job." When proposing changes, students prioritized surface-level issues such as grammar or mechanics and paid little attention to macro issues. The few students who made comments on content expressed them in inappropriate ways by being too negative and not giving further suggestions on how to improve the essay. When revising, students did not always make many changes, either because they did not trust the validity of the peers' feedback or because they did not receive good quality comments on their papers.

After analyzing and reflecting on the rather limited results, we concluded that the training was too basic and did not help students develop a deep understanding of peer review. The training consisted of: (1) explaining the assignment and reminding about the deadlines; (2) justifying the importance of peer review, in part by summarizing the research done in the field; (3) giving students a paper copy of the peer guidelines, which they were expected to follow during the peer review; (4) showing the feedback taxonomy on a Power Point slide and briefly explaining the different types of feedback; and (5) giving students a handout with instructions on how to use the peer review software. In this first round of training, students were not given good examples of high-quality peer review, nor did they have the chance to practice commenting before completing the peer review activity. In addition, class lessons and the content of instruction were disconnected from the peer review activity. This made the training even more distant and unfamiliar to the students. Overall, the peer review training was inappropriate, or at least not sufficient, to yield the beneficial results often associated with this pedagogical activity.

Analysis and Reflections on the Second Implementation

For the second implementation of the project, Summer 2017, we thought the solution was to present many activities, as it is shown in Hu (2005). Possible activities were separated into five groups according to their function, and we decided to increase the amount of activities in each of

the groups. As shown in Table 2, 10 new activities were added, presenting a total of 16 activities. The first group included two basic activities. The first activity was presenting the project to the students and giving them an overall picture of the main tasks to be completed in a period of four weeks. At this early stage of the project, the instructor showed adapted examples from former students so that they had a clearer idea of what their performance should look like at the end of the project. The second activity was ensuring a direct alignment between the class sessions and the requirements of the peer review activities. Students learned about personal narratives from the lens of peer reviewing, always keeping in mind the specific dimensions of the guidelines. For example, by deconstructing and analyzing exemplary draft models in terms of their dimensions (narrative elements), students not only became familiar with the guidelines, but also learned the characteristics of a good personal narrative.

The second group of activities aimed to raise awareness about a series of issues, including the importance of peer review, the recent research findings in the field, the extended use of peer review in journal and academic presses, the advantages and challenges of using the methodology, and additional difficulties that result from applying peer review in L2 learning contexts. These activities were carried out in a small group format, then shared as a class discussion, and when necessary, briefly repeated or summarized by the instructor. As shown in previous studies (e.g. Berg, 1999b; Hu, 2005; Min, 2005; Paulus, 1999; Stanley, 1992), awareness-raising is an essential step to create a positive class atmosphere, engage students in the peer activity, and guide them towards developing appropriate attitudes as both writers and readers.

The session dedicated to awareness-raising was followed by activities related to the peer review guidelines. Students were encouraged to use the peer review guidelines as a potential solution to counteract some of the disadvantages pointed out in the discussions. After giving the students a copy of the guidelines, there were two additional discussion-based activities. The first focused on the role of the guidelines in the review process. The second encouraged learners to analyze the questions in terms of their format and their foci of attention (e.g., global issues were prioritized over surface issues, and all questions were open-ended as to avoid yes/no responses).

To ensure that students understood each criterion of the guidelines, they read all the questions from top to bottom in small groups. For every question, they discussed confusing parts and verified that all questions were logical, understandable, and specific enough. Then, we used

a written sample to briefly model how to use the guidelines. The modeling emphasized four stages: (1) reading the paper from top to bottom to understand the overall narrative; (2) attending to macro issues from the first section of the guidelines; (3) attending to micro issues from the second section of the guidelines; and (4) reviewing all comments before submitting. While these stages may be somewhat general and basic, Hu (2005) argued they help counteract students' tendencies to focus on surface issues only.

The fourth group of training focused on the feedback comments, and in particular the quality of their content. Motivated by studies that focus on feedback typology (Min, 2005; Nelson & Schunn, 2009; Stanley, 1992), students were encouraged to give feedback comments that were rich, varied, and thoughtfully constructed. To emphasize this information, two main activities were presented in class. The first activity involved showing and explaining the different types of feedback available to them: positive comments, problem identification, suggestion, alteration, justification, elaboration (typology adapted from Lu & Law, 2012, Liu & Sadler, 2003, and Nelson & Schunn, 2009). Then, according to Min's (2005) suggestion, the instructor modeled how to use this typology by following a five-step procedure: (1) understanding the writers' intentions; (2) localizing the problem; (3) explaining in detail the nature of this problem; (4) giving specific suggestions on how to solve or improve the issue; and (5) justifying how the suggestions may help improve the overall text. As the researcher affirmed, this step-by-step guidance helps make students feedback "more relevant and specific because it revolv[es] around clarifying, identifying and expounding on a single issue and suggesting a way to improve it" (Min, 2005, p. 303).

The last series of activities involved two technology training tasks. First, students read the handout that clearly explained the step-by-step procedures on how to use the peer review software (Turnitin Feedback Studio). This handout was available to them both in paper and online. Second, the instructor demonstrated in class the main uses of the software and how to get started with peer review. Students were allowed to use their laptops to follow the steps projected on the screen. The goal of this training was to offer students a basic understanding of the software and to facilitate the reviewing activity at home.

Overall, we found the training to be more complete and more successful than the one provided the previous semester (Spring 2017). Students receiving the training in Summer 2017

gave more feedback and of better quality on their classmates' writing. Their comments were generally longer, more varied, and revision oriented (as opposed to affective comments). Their attitudes toward peer review, as reported in the individual informal conferences, also improved. Reading other peers' work was interesting, and they saw that their role as reviewers was to be helpful and give their opinion from a reader's perspective.

The end-of-the-semester questionnaire guided which improvements to make for the next session. Based on student responses (i.e., what they liked the most and the least), we learned that students enjoyed the project, the overall pace of the assignments, and the anonymity and convenience of using online software. In addition, a theme in students' responses was that the activities were "too long and boring" and suggested that future activities be "shorter" and "more dynamic."

Based on analysis and reflections, it appears that students had developed positive attitudes towards peer review. However, there was still a lot of room for improvement in instructional practices. With the students' perspective in mind, we reviewed the activities presented in the project and realized that they were often too theoretical and did not always engage students to be active. With the exception of the small and large group discussions, there were not many opportunities for actual practice. After coming to this realization, and inspired by the positive results of the project, we focused on making the explanations more practical. We planned activities that were more dynamic and engaging, and that required students to apply their knowledge. Below are the changes made for Fall 2017.

Analysis and Reflections on the Third Implementation

Based on the analysis and reflections from the previous groups, we retained 14 of the 16 activities. In each group we added one to three activities that were hands-on and required students to apply their learning. Seven new activities were added to the list for a total of 21 activities making up the peer review training.

One new activity pertained to the presentation portion of the project. Students were asked to bring at least two pictures about two personal experiences, discuss the pictures with peers, and build their narrative based on the picture selected. The purpose was to make the

overall project more engaging and to promote more interesting peer review practices in which students could connect the text with the picture.

Another activity added was in the group of awareness-raising activities. After the discussion session on the importance and usefulness of peer review, students received a list of ground rules, principles, and attitudes that work best in peer review (adapted from The Teaching Center at Washington University, 2009). The activity started at home, where they had to mark with a star the five rules or principles they thought were most important. They shared their responses and decided as a class the most essential rules and principles for peer review. The result was a ten-item list serving as the code of conduct for everyone to respect. This code of conduct helped create a comfortable and open working atmosphere among students.

A third activity added was in the category for guidelines. Through a set of hands-on activities, students were encouraged to practice with the guidelines and apply their knowledge. The activities included short matching activities (e.g., linking goals with questions, the two formats included in the guidelines as presented in Table 1), transforming inadequate and rude comments into appropriate, courteous comments, and classifying good comments according to the dimension of the guidelines they thought it related to. These activities were performed in small groups and then shared with the whole class. The main purpose was to actively engage students in the use and understanding of the peer review guidelines while also boosting their familiarity with the various dimensions included in the guidelines.

The category of feedback comments required the greatest change. In previous groups it had not been enough to explain to students the types of feedback and modeling how to apply them. In fact, some did not enjoy the modeling part, or they simply disconnected from the activity. Therefore, we decided to substitute the modeling activity for three other activities that required active participation from the students. In the first activity, students were given a set of feedback comments to classify by type. The purpose was for students to acknowledge the variety of feedback available to them. The second activity went a step further as students were asked to produce rich and varied comments in response to hypothetical situations (e.g., imagine a plot that enumerates a series of actions without describing a major scene). To help students formulate longer and more complex comments, the instructor guided students through the step-by-step procedure suggested in Min (2005): (1) understand the writers' intentions; (2) localize the

problem; (3) explain the problem; (4) offer a specific suggestion; and (5) justify the suggestion. The last activity in this group involved practicing peer review with actual samples from former students. Several sessions were spent on this group of activities, from practicing with excerpts that related directly to a specific dimension of the guidelines, to working with full texts and all the dimensions in the guidelines. The activities were completed in small groups first, then shared and discussed with the class. Topics related to feedback variety, specificity, appropriacy, and insightfulness were considered in the discussions.

The activities for practicing peer review continued with the fifth group, technology training. Students were asked to respond to a draft written by a former student, but this time they typed the comments into Turnitin Feedback Studio. The reviewing activity was completed in small groups and then shared as a whole class.

The main goal of the technology training was to simulate the peer review context students would find themselves in when they did peer review on their own. Moreover, we wanted all possible technology issues to arise before students tried to complete the activity at home. Thus, students were encouraged to try everything in the software, including logging out to continue later with their work. If anyone ran into a problem, the instructor not only assisted the student individually, but also shared the problem with the rest of the class so they knew how to solve it if it happened to them. It became apparent by the third semester that this session was highly important to reduce the number of problems students would have at home, facilitate the work of the reviewers, and raise their confidence in giving online comments.

Conclusions

After implementing an L2 writing project for three consecutive semesters, the project developed smoothly and with positive results by the third implementation (Fall 2017). Students in this group successfully met the goals and expectations of the project. After incorporating dynamic and hands-on activities, removing modeling activities, and shortening theoretical tasks, the group of Fall 2017 demonstrated themselves to be the most engaged and with the most positive attitude toward peer review. They also understood best the responsibilities of peer review, and appropriately followed the procedures required for an efficient review.

In terms of comments given on peers' writing, students strove to give high-quality feedback that not only praised their peers but also guided them to improve their essay. They gave critical and constructive comments and were able to follow the five-step procedure that led students to analyze, locate and explain a problem, and offer a suggestion and justify the solution (Min, 2005). While some of the revision-oriented comments pointed out issues that were not necessarily problematic, they still called the writers to reconsider their text and work toward improving and polishing their narratives. These types of comments are equally valid and enriching when complemented by other types of feedback (e.g., problem identification, positive feedback) and contributed to giving writers more varied and rich feedback.

In addition to macro- and meaning-related comments, the students also commented on surface issues related to grammar, vocabulary choice, and use of punctuation. However, these types of comments were less abundant than the macro-level comments because they received less priority during peer review. Taken together, both the micro and macro types of comments served as a valuable source of information to the writers who needed to revise their draft. Students showed that they took peer review seriously, wanting to help their peers by providing detailed and well thought out feedback.

During informal conferences, students showed that they valued the overall project. They reported how both the training and the reviewing activities helped them become more critical not only of peers' papers but also of their own paper, ultimately helping them become better writers. Similarly, they explained that the high-quality comments received helped improve their papers, as overlooked problematic issues were pointed out. These comments also helped the writer develop a sense of audience, as they realized how the readers' perspective often lacks the background information privy to the author.

Many students insisted in the end-of-the-year questionnaire that the structure and organization of the project should be retained. Many felt that the guidance they received from the training helped them become better critical thinkers and provided them with the necessary tools to give better quality comments to their peers. More importantly, students generally enjoyed knowing what to do and what was expected of them at all times, since they had previously practiced all the steps in class.

Overall, the results are similar to Hu (2005) in that after a recursive process of reflecting and modifying teaching practices each semester, the last implementation lead to more positive results. However, the procedures we followed in the L2 writing project design and the peer review activities differed from those proposed in Hu (2005). First, the L2 writing project and the peer review training were designed to meet the requirements of the online peer review setting, since Hu's project design (2005) was for a face-to-face peer review context. Second, the activities that constituted the peer review training were created to respond to the specific class environment and the profile of the students in this study. In particular, the activities proposed in the third implementation were more dynamic and hands-on, encouraging students to actively participate and take ownership of their learning. Conversely, activities in which learners had to take a passive attitude, such as modeling and instructing, were eliminated or at least shortened. It is possible that the Chinese context in Hu's study (2005) allowed learners to benefit from more passive activities involving listening and just paying attention. In the U.S. college context however, these activities resonated the least among the students, whereas other types of activities, such as those added in the last implementation (Spring 2017), appeared to be most successful.

While these changes present additional benefits to the project, it is important that other researchers join the cause of improving the pedagogy of peer review practices. There is much to be explored about peer review and how training may maximize students' critical skills, the quality of their feedback, their written performance, and the development of their L2 writing skills.

References

- Allen, D., & Mills, A. (2016). The impact of second language proficiency in dyadic peer feedback. *Language Teaching Research*, 20(4), 498–513.
- Berg, E. C. (1999a). Preparing ESL students for peer response. *TESOL Journal*, 8(2), 20–25.
- Berg, E. C. (1999b). The effects of trained peer response on ESL students' revision types and writing quality. *Journal of Second Language Writing*, 8(3), 215–241.
- Carr, W., & Kemmis, S. (2003). *Becoming critical: Education knowledge and action research*. New York, NY: Taylor & Francis.
- Carson, J. G., & Nelson, G. L. (1996). Chinese students' perceptions of ESL peer response group interaction. *Journal of Second Language Writing*, 5(1), 1–19.
- Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction*, 20(4), 328–338.
- Cho, Y. H., & Cho, K. (2011). Peer reviewers learn from giving comments. *Instructional Science*, 39(5), 629–643.
- Connor, U., & Asenavage, K. (1994). Peer response groups in ESL writing classes: How much impact on revision? *Journal of Second Language Writing*, 3(3), 257–276.
- Couzijn, M., & Rijlaarsdam, G. (2005). Learning to write instructive texts by reader observation and written feedback. In G. Rijlaarsdam, H. Bergh, & M. Couzijn (Eds.), *Effective learning and teaching of writing* (pp. 209–240). New York, NY: Kluwer.
- de Guerrero, M., & Villamil, O. S. (2000). Activating the ZPD: Mutual scaffolding in L2 peer revision. *Modern Language Journal*, 84(1), 51–68.
- Dunlap, J. C. (2005). Workload reduction in online courses: Getting some shuteye. *Performance Improvement*, 44(5), 18–25.
- Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction*, 20(4), 304–315.
- Ho, M., & Savignon, S. J. (2007). Face-to-face and computer-mediated peer review in EFL writing. *CALICO Journal*, 24(2), 269–290.
- Hu, G. (2005). Using peer review with Chinese ESL student writers. *Language Teaching Research*, 9(3), 321–342.
- Hyland, F. (2000). ESL writers and feedback: Giving more autonomy to students. *Language Teaching Research*, 4(1), 33–54.
- Hyland, K., & Hyland, F. (2006a). Feedback on second language students' writing. *Language Teaching*, 39(02), 83–101.
- Hyland, K., & Hyland, F. (2006b). Interpersonal aspects of response: Constructing and interpreting teacher written feedback. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 206–224). New York, NY: Cambridge University Press.

- Kamimura, T. (2006). Effects of peer feedback on EFL student writers at different levels of English proficiency: A Japanese context. *TESL Canada Journal*, 23(2), 12–39.
- Lam, R. (2010). A peer review training workshop: Coaching students to give and evaluate peer feedback. *TESL Canada Journal*, 27(2), 114–127.
- Leki, I. (1990). Potential problems with peer responding in ESL writing classes. *CATESOL Journal*, 3(1), 5–19.
- Levi Altstaedter, L. (2016). Investigating the impact of peer feedback in foreign language writing. *Innovation in Language Learning and Teaching*, 12(2), 137–151.
- Liu, J., & Hansen, J. G. (2002). *Peer response in second language writing classrooms*. Ann Arbor, MI: University of Michigan Press.
- Liu, J., & Sadler, R. W. (2003). The effect and affect of peer review in electronic versus traditional modes on L2 writing. *Journal of English for Academic Purposes*, 2(3), 193–227.
- Lu, J., & Law, N. (2012). Online peer assessment: Effects of cognitive and affective feedback. *Instructional Science*, 40(2), 257–275.
- Mendonça, C. O., & Johnson, K. E. (1994). Peer review negotiations: Revision activities in ESL writing instruction. *TESOL Quarterly*, 28(4), 745–769.
- Min, H. (2005). Training students to become successful peer reviewers. *System*, 33(2), 293–308.
- Min, H. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing*, 15(2), 118–141.
- Min, H. (2008). Reviewer stances and writer perceptions in EFL peer review training. *English for Specific Purposes*, 27(3), 285–305.
- Nasrollahi, M. A., Krish, P., & Noor, N. M. (2012). Action research in language learning. *Procedia-Social and Behavioral Sciences*, 47, 1874–1879.
- Nelson, G. L., & Murphy, J. M. (1992). An L2 writing group: Task and social dimensions. *Journal of Second Language Writing*, 1(3), 171–193.
- Nelson, G. L., & Murphy, J. M. (1993). Peer response groups: Do L2 writers use peer comments in revising their drafts? *TESOL Quarterly*, 27(1), 135–141.
- Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science*, 37(4), 375–401.
- Nilson, L. B. (2003). Improving student peer feedback. *College Teaching*, 51(1), 34–38.
- Nystrand, M. (1986). *The structure of written communication: Studies in reciprocity between writers and readers*. Orlando, FL: Academic Press.
- Papadopoulos, P. M., Lagkas, T. D., & Demetriadis, S. N. (2012). How to improve the peer review method: Free-selection vs assigned-pair protocol evaluated in a computer networking course. *Computers & Education*, 59(2), 182–195.
- Paulus, T. M. (1999). The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing*, 8(3), 265–289.

- Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT Journal*, 59(1), 23–30.
- Stanley, J. (1992). Coaching student writers to be effective peer evaluators. *Journal of Second Language Writing*, 1(3), 217-233.
- Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of Second Language Writing*, 14(3), 153–173.
- Strijbos, J., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction*, 20(4), 291–303.
- The Teaching Center at Washington University. (2009). *Commenting on student writing*. Retrieved from <https://teachingcenter.wustl.edu/resources/writing-assignments-feedback/commenting-on-student-writing/>
- Tsui, A. B., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing*, 9(2), 147–170.
- Vorobel, O., & Kim, D. (2014). Focusing on content: Discourse in L2 peer review groups. *TESOL Journal*, 5(4), 698–720.
- Wigglesworth, G., & Storch, N. (2012). What role for collaboration in writing and writing feedback. *Journal of Second Language Writing*, 21(4), 364–374.
- Yang, M., Badger, R., & Yu, Z. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing*, 15(3), 179–200.
- Zhang, S. (1995). Reexamining the affective advantage of peer feedback in the ESL writing class. *Journal of Second Language Writing*, 4(3), 209–222.
- Zhao, H. (2010). Investigating learners' use and understanding of peer and teacher feedback on writing: A comparative study in a Chinese English writing classroom. *Assessing Writing*, 15(1), 3–17.
- Zhu, W. (1995). Effects of training for peer response on students' comments and interaction. *Written Communication*, 12(4), 492–528.

Investigating an L2 French Writer's Performance on an Integrated Task

Anna Evans, University of Iowa

This study presents a research-in-progress report on a larger dissertation research project. It offers a definition of integrated writing tasks and highlights their benefits for foreign language education. It provides an overview of relevant research studies on writing processes, written products, and French writing as related to the use of sources in second language learners' essays. The aim of the corresponding empirical study was to investigate what writing processes a French learner engaged in and how she integrated sources (article, graph, and audio recording) into her essay while completing an integrated writing task. The rationale and methodology of the study are described in detail and preliminary results based on the pilot data are discussed.

Traditionally, writing assessment in foreign language (FL) classes is based on independent tasks, which require second language (L2) learners to compose essays in response to a short prompt. However, independent writing tasks are often critiqued for being inauthentic in academic contexts and for providing little insight into the full range of L2 learners' writing abilities (Weigle, 2004). Concerns with the limitations of independent writing tasks have resulted in increased interest in integrated writing tasks among L2 researchers, test developers, and practitioners (Plakans, 2015; Yu, 2013). Integrated writing tasks are defined as

tasks in which test takers are presented with one or more language-rich source texts and are required to produce written compositions that require (1) mining the source texts for ideas, (2) selecting ideas, (3) synthesizing ideas from one or more source texts, (4) transforming the language used in the input, (5) organizing ideas, and (6) using stylistic conventions such as connecting ideas and acknowledging sources. (Knoch & Sitajalabhorn, 2013, p. 306)

Indeed, most academic course-based tasks are integrated writing tasks. Students are rarely required to produce academic writing without any references, but rather expected to summarize, synthesize, and respond to sources in their writing (Cumming, 2014; Weigle, 2004). Since integrated writing tasks resemble writing for academic purposes much more closely than independent writing tasks, they are also worthwhile for FL education and should be more widely incorporated into upper-level FL classes.

Due to the relative novelty of integrated writing tasks in L2 contexts, scholars (Cumming, 2014; Knoch & Sitajalabhorn, 2013; Yu, 2013) have called for more rigorous and systematic research in the area. A thorough review of the existing research literature on integrated writing tasks has revealed a number of significant research gaps. First, scholars who investigate integrated writing tasks tend to focus on either writing processes or written products and investigate them separately. However, to gain deeper insight into L2 learners' performances on integrated writing tasks, writing processes and written products require further investigation together rather than separately. In the modern computerized era, there is also a need to examine what writing processes L2 learners engage in on a computer and what their written products look like as a result. Second, the existing research on integrated writing tasks has been primarily conducted in English as a second/foreign language contexts and should be rather expanded to include other languages. Finally, a limited number of studies have investigated how L2 learners engage with modalities (audio and visual sources) other than print sources. This study addresses the identified research gaps by investigating an L2 French learner's writing processes and written product based on an integrated writing task that contains print, audio, and visual sources.

Literature Review

Writing Process Research

Writing processes have been extensively studied in first language (L1) and L2 writing literature; however, fewer researchers (Flower, 1994; Hayes, 1996) have accounted for the role of reading in writing processes. Prominently, Spivey (1984) referred to the processes of discourse synthesis—selecting, connecting, and organizing—to demonstrate unique characteristics of L1 source-based writing. Discourse synthesis is thought to be able to establish links between reading, listening, and writing skills, since the processes of connecting and organizing are both components of reading and listening comprehension and of writing processes. From Spivey's research, we know that L1 source-based writing processes are considerably different from L1 writing processes without the use of sources. Since writing processes are always cognitive processes (Flower & Hayes, 1981), L1 source-based writing is expected to provoke different cognitive processes as well. The situation becomes even more complicated in L2 source-based writing. Significantly, L2 writing processes are different from L1 writing processes because they

involve thinking in two languages and moving between two cultural and rhetorical dimensions (Hirvela & Du, 2013).

A few researchers (Allen, 2004; Li & Casanave, 2012; Ruiz-Funes, 1999) have investigated L2 source-based writing processes. Of special relevance to this study is Plakans's (2007) dissertation study in which she analyzed L2 English students' think-aloud protocols and proposed a model for writing processes in response to a reading-writing argumentative task. Plakans divided the writing processes into two main stages: preparing to write and writing. While preparing to write, participants engaged in a rather linear sequence of activities. They started by reading the task prompt, instructions, and source texts, then tried to understand the task and expressed their opinions, and, finally, planned and organized the content of their essays. While writing, participants engaged in recursive writing processes. They wrote, planned, rehearsed wording, used the source texts, reread and evaluated their own writing. Plakans also found that interested writers used the processes of discourse synthesis: While using the source texts, they reread, selected, paraphrased, and cited more than other writers.

Written Product Research

What considerably distinguishes integrated writing tasks from independent tasks is the integration of sources into written products. Among earlier studies, Cumming, Kantor, Baba, Erdosy, Eouanzoui, and James (2005) studied the relationship between 36 L2 English learners' source use across task types (independent, reading-writing, and listening-writing) and proficiency levels. Higher-scoring writers tended to synthesize or summarize information from the sources in a coherent manner. Mid-range proficiency learners tended to paraphrase or copy verbatim from the sources. Lower-scoring participants did not use much information from the sources and often wrote their essays based on their personal experience, as they might not have comprehended information in the sources.

On a larger scale, Plakans and Gebiril (2013) further investigated how source use predicted 480 L2 English writers' scores on a reading-listening-writing task. They found that as scores went up, the use of ideas from the reading passage and verbatim source use decreased, whereas the use of ideas from the lecture, the inclusion of important ideas from the sources, and

implicit source use (paraphrasing) increased. Their research is particularly pertinent to this study because of the inclusion of both print and audio sources in their analysis.

L2 French Writing Research

In contrast to L2 English writing research described above, L2 French writing is relatively understudied (Benevento & Storch, 2011), and studies on L2 French source-based writing tasks (Corbeil, 2000; Cumming, Rebuffort, & Ledwell, 1989; Rivard, 2001) are very few in number. Whereas Cumming et al. (1989) explored issues other than source use, Rivard (2000) touched upon source use in L2 French summaries. In particular, he analyzed both language and content in L1 and L2 French high school students' summaries. He found that in the final grade level, L1 and L2 French students' summaries differed significantly only in style, with L1 French students being superior. However, none of these studies on L2 French source-based writing tasks has addressed the issues of both L2 French learners' writing processes and written products in terms of their source use.

Further studies on integrated writing tasks are warranted so as to contribute to the existing body of research. Future studies should examine both writing processes and written products in terms of source use and expand to include languages other than English. In addition, it would be revealing to investigate how L2 learners compose essays in response to not only print, but also audio and visual sources. Investigating integrated writing tasks across different contexts, languages, and modalities will result in a better understanding of learners' performances on this writing task, which is still relatively new in L2 contexts.

Research Questions

Based on the identified research gaps in the existing research literature, I address the following research questions in this study:

1. What writing processes does an L2 French writer engage in while completing an integrated writing task?
2. What does an L2 French writer integrate from sources into her written product?

Methodology

It is important to mention that this study is a research-in-progress report on the design and progress of a larger dissertation research project. This study includes data from the corresponding pilot study, and the results presented here are preliminary and exploratory.

Participant

In the spring 2018 semester, I recruited a female student named Kim (pseudonym) to pilot test data collection instruments and procedures. At the time of the pilot testing procedures, Kim was 21 years old and a third-year student majoring in French and linguistics at a large public Midwestern university. Her native language is English. Kim studied French for three years in high school and continued taking French courses at the university. She had taken five upper-level French courses by the time of the pilot testing procedures. Kim stated that she enjoyed writing in English and French: “It helps me practice my writing and reading skills.” Kim shares many characteristics (age, native language, university student status, enrollment in upper-level French courses) with students who later participated in the actual data collection procedures. In the actual data collection, I recruited 38 students to complete an integrated writing task and questionnaires. Eight of them volunteered to do a think-aloud protocol while writing their essays and were then interviewed within three days.

Data Collection Procedures

At the beginning of the pilot testing procedures, Kim completed a brief background information questionnaire. I then trained her how to think aloud by showing an example video of a person thinking aloud and by offering her a practice writing activity. After the think-aloud training, Kim received an integrated writing task. First, Kim studied the instructions, the prompt, and the three sources (article, graph, and audio recording) that presented different views on the topic and then wrote an argumentative essay. The integrated writing task was taken from the Advanced Placement French Language and Culture Exam Free-Response Questions (College Board, 2013) with permission from the College Board. The prompt was, “*Faut-il arrêter de manger de la viande et devenir végétarien pour être en bonne santé?*” [Should we stop eating meat and become vegetarian in order to be healthy?].

While writing her essay in response to the integrated writing task, Kim did a think-aloud protocol. I used the screen-capture software program Panopto to record video of the computer screen as well as audio of Kim's voice. Many of the studies on source-based writing processes (Hirvela & Du, 2013; McCulloch, 2013; Plakans, 2007; Yang & Shi, 2003) employed L2 writers' think-aloud protocols. Although think-aloud protocols have been critiqued for possibly disrupting and altering participants' thinking processes (McCulloch, 2013), they still provide important insights into participants' writing processes, intentions, motivations, and decision-making processes.

Having finished her essay, Kim completed a retrospective source integration questionnaire that contained questions about her perceptions of the integrated writing task. This questionnaire was adapted with permission from a reading-to-write process questionnaire designed by Gebril and Plakans (2009) and modified to include more questions about the audio and visual sources. At the end of the pilot testing procedures, I interviewed Kim to clarify her responses on both questionnaires and to learn about her perceptions of the think-aloud effect and the integrated writing task.

Data Analysis

To answer the first research question about Kim's writing processes, I analyzed Kim's screen recording in conjunction with her think-aloud protocol. I transcribed, searched for, coded, and categorized Kim's verbal accounts of her writing processes. Although the initial codes were based on Plakans's (2007) model for source-based writing processes, the coding process was still primarily inductive and was modified to reflect Kim's source use and writing processes. I also analyzed Kim's responses on both questionnaires and in the interview. Her responses provided additional data to support the analysis of the think-aloud protocol as well as identify and solve issues with the data collection instruments and procedures for the actual data collection.

To answer the second research question about Kim's written product, I followed Shi's (2004) and Weigle and Parker's (2012) procedures to analyze what Kim integrated from the sources into her essay. First, I identified borrowed strings, which Weigle and Parker define as "a combination of two or more content words (nouns, verbs, adjectives, adverbs) from the same clause within the source text with or without additional intervening words, or a string of at least

three consecutive words that form a syntactic constituent” (pp. 121–122). I identified borrowed strings as referenced, non-referenced, and quotations. I further categorized referenced and non-referenced borrowed strings as exactly copied, slightly modified, and substantially reformulated. Finally, I calculated the percentages of total textual borrowing, per category (referenced and non-referenced), and per subcategory (exact copies, slight modifications, and substantial reformulations).

Results

Research Question One

While preparing to write her essay, Kim mostly tried to comprehend the sources by interpreting the graph and rereading and translating the article. While interacting with the sources, Kim also expresses her own feelings (“Yikes!”) and opinions on the topic, often disagreeing with the author of the article and questioning the trustworthiness of the information presented, e.g., “He also says that fish is less rich in iron which does not have a lot of risk, which isn’t very true, but that’s fine.” Kim then looks through her notes on the audio recording and translates them and comments on the trustworthiness of the interviewee’s statements. Finally, she summarizes the main idea of the article and expresses her opinion:

So, this reading passage is saying to only eat one or two times a week and then supplement that with fish. So, you are really eating the same amount of animals, you are just supplementing it with a different animal, which really doesn’t make any sense.

As is evident from the overview of Kim’s pre-writing stage, she actively engages with the sources and evaluates them critically, and her approach is in line with the argumentative writing task which requires her to integrate information from all the sources and state and defend her own opinion on the topic.

While writing her essay, Kim provides her own arguments and relies primarily on the ideas from the audio recording in the first three paragraphs of her essay because these ideas align more closely with her own views on the topic. Similar to her reading strategies in the pre-writing stage, Kim engages in the processes of writing and translating her own writing, “*Je pense que plus des gens doivent être végétariens* [I think that more people should be vegetarians], because it’s good

for the environment, because vegetarianism is good for the environment, *parce que végétarianism est bien pour l'environnement* [because vegetarianism is good for the environment].”

Having completed half of the essay, Kim pauses and scrolls through the task and rereads the task instructions to verify whether her essay meets the task requirements. After that, she rereads what she has written so far and decides to better organize her essay and to strengthen her introduction by adding one more sentence that combines ideas from the article and the audio recording. In the middle of her writing process, she continues rereading her essay again and again, makes multiple revisions, erases some parts of her essay, and works on the organization of her essay by dividing her essay into paragraphs.

Having made some substantial revisions, Kim starts actively using ideas presented in the article. She rereads the article, checks the author’s name, cites his name in her essay and incorporates several ideas from the article into her essay. However, Kim seems to have trouble integrating ideas from the article with her own arguments in a coherent manner. Indeed, Kim struggles with one important part of argumentative writing—using a source that presents a point of view that opposes her own. In the end, she responds to the ideas from the article by either rejecting them or by rephrasing them to rather reflect her own views on the topic. Overall, at this stage, Kim engages in the same strategies of rereading and translating aloud the article and her essay, and she also occasionally comments on the article and expresses her opinions on the topic.

Before writing her conclusion, Kim rereads her essay from the beginning once again and makes several revisions. She then writes the final sixth paragraph that reflects her views on the topic and reiterates the ideas stated in the introduction. Finally, Kim rereads her essay from the beginning one more time and makes several revisions before submitting the essay.

The main writing processes that Kim engaged in while completing the integrated writing task were: (1) writing and translating her own essay; (2) rereading and translating the article, her notes on the audio recording, and her own essay; (3) selecting ideas from the sources; (4) commenting on the task, the sources, and the topic and expressing her feelings; (5) organizing her essay; (6) rereading and revising her essay; and, to a lesser extent, (7) summarizing the main ideas from the sources. As is seen from the analysis of Kim’s think-aloud protocol, she often translates the sources and her own essay. However, in the interview, Kim said that, “I like to think that I do it [translate] minimally, just because it’s easier to think in French if you are

writing in French.” It is pertinent to note that this discrepancy between the analysis of the think-aloud protocol and Kim’s perception of her reading and writing processes may be due to the requirements of the think-aloud procedure; that is, Kim was asked to say out loud everything that was going through her mind and she was allowed to use both English and French in her think-aloud protocol.

Research Question Two

Kim’s essay contained 329 words, and 93 of those words (28% of the whole essay) were found to be instances of textual borrowing from the article and the audio recording. Most of Kim’s textual borrowing was referenced (58% of her total textual borrowing), i.e., attributed to the author of the article or to the interviewee in the audio recording. Kim borrowed substantially more from the article than from the audio recording: She borrowed 65 words from the article (70% of her total textual borrowing) and 28 words from the audio recording (30% of her total textual borrowing). Such a substantial difference can be explained by the fact that Kim had access to the article the whole time while writing her essay. In contrast, she did not have access to the audio recording after having listened to it twice and taken some notes.

A more detailed analysis showed that Kim used 27 words (29% of her total textual borrowing) as quotations from the article. For instance, Kim wrote (original spelling and punctuation preserved), “*Alain Soussa a dit que, ‘C’est la conclusion de l’une des plus importantes études réalisées sur les liens entre consommation de viande et cancer du colon’*” [Alain Soussa said that, ‘This is the conclusion of one of the most important studies conducted on the links between meat consumption and colon cancer’]. Here, Kim copied the original text (in bold), used quotation marks, and cited the name of the author of the article.

Thirty words (32% of her total textual borrowing) were substantial reformulations from both the article and the audio recording. For example, Kim wrote, “*C’est évident que la viande n’est pas bien pour notre santé, on peut prendre les **protéins**, le **fer**, et les **vitamins** d’autre alimentaires pas la viande*” [It is evident that meat is not good for our health, one can get proteins, iron, and vitamins from other food not meat]. Here, Kim substantially reformulated the original text (in bold) but did not cite the author of the article.

Thirty-six of the words (39% of her total textual borrowing) were exact copies from the article and the audio recording. The inclusion of exact copies in a student's essay is typically of most concern to L2 instructors. However, a closer analysis revealed that 21 (in bold) out of those 36 words were in the same sentence, "*Alain Soussa a dit que si tu ne veut pas manger végétarien, tu peut **diminuer ta consommation de viande rouge une à deux fois par semaine maximum et augmenter la part de volailles et de poisson***" [Alain Soussa said that if you do not want to eat vegetarian, you can diminish your consumption of red meat one to two times per week maximum and increase the amount of poultry and fish]. Although Kim attributed the idea to the author of the article, she did not use conventional quotation marks to identify the borrowed string as she had done earlier.

Even though Kim included more information from the article (70% of her total textual borrowing) than from the audio recording in her essay, she still incorporated a substantial amount of content from the audio recording. She used the information presented in the audio recording to support her own opinion on the topic since the views expressed in the audio recording aligned closely with her own. Most of the instances of textual borrowing (58% of her total textual borrowing) were referenced, so Kim seems to be aware of the appropriate citation practices and conventions of source use in her own writing. Paraphrasing posed an apparent challenge for Kim, as most of the instances of her textual borrowing were either quotations or exact copies whereas substantial reformulations represented less than a third (32%) of total textual borrowing in her essay. It should be noted that I refrain from judging whether the amount of Kim's source use in her essay is acceptable. If Kim had written this essay for one of her French courses, her instructor would judge whether the amount of textual borrowing in Kim's essay was appropriate and met their expectations.

Discussion and Conclusion

Following Plakans's (2007) model, Kim's writing processes were also divided into two main stages: preparing to write and writing. While preparing to write her essay, Kim engaged in rereading and translating the sources, mostly the article. She also positioned herself in relation to the task and the topic by commenting on the topic and expressing her own opinion. While

writing her essay, Kim, like participants in Plakans's study, engaged in recursive writing processes. In particular, Kim wrote, translated, used the sources, reread and revised her essay.

Additionally, Kim was interested in the topic of the integrated writing task, which she stated in her think-aloud protocol, in the questionnaire, and in the interview. According to Plakans (2007), Kim can be described as an interested writer. Plakans found that interested writers engaged in more constructive or problem-solving processes, interacted more with the sources, and thought more about their opinions than the other writers. While using the sources, Kim—like the interested writers in Plakans's study who reread, selected, paraphrased, and cited more than the other writers—used the processes of discourse synthesis.

In terms of Kim's essay, most of the language that Kim integrated from the sources into her essay came from the article, whereas less language came from the audio recording and none from the graph. However, one third of her total textual borrowing was from the audio recording, which is still a substantial amount. This demonstrates that Kim comprehended well the audio recording and managed to effectively use language from the audio recording to support her own opinion on the topic. Significantly, Plakans and Gebril (2013) found that as scores went up, the use of ideas from the reading passage and verbatim source use decreased, whereas the use of ideas from the lecture, the inclusion of important ideas from the sources, and implicit source use (paraphrasing) increased. On the one hand, Kim comprehended well the audio recording and integrated important ideas from both the article and the audio recording into her essay. On the other hand, she struggled with paraphrasing and often resorted to verbatim source use. The fact that Kim's approach to the integrated writing task differed from that of test takers in Plakans and Gebril's study could at least partially be explained by differences in the task instructions and requirements (summary vs. argumentative essay).

Overall, analyzing both writing processes and written products helps identify L2 learners' difficulties with completing writing tasks and then select appropriate pedagogical approaches to address those difficulties. Without the analysis of Kim's writing processes, we would not know the reasons for weaknesses in Kim's essay. For instance, several sentences in her essay appear to misrepresent ideas from the article, possibly due to lack of comprehension. However, thanks to Kim's think-aloud protocol, we know that she understood well the ideas in the sources. Rather, the issue lies in Kim's difficulty using a source that contradicts her opinion. Additionally, thanks

to Kim's think-aloud protocol, we know that she did not make an outline before she started writing. The absence of a thought-through outline might have intensified her trouble with integrating ideas from the article with her own arguments in a coherent manner. This analysis results in a better understanding of what kind of pedagogical intervention Kim needs in the future and what should be taken into consideration in terms of subsequent L2 writing instruction.

Finally, the pilot testing allowed me to detect Kim's misunderstandings of the instructions, problems with the data collection instruments and procedures, and technical issues with the software programs. Thanks to the findings made during the pilot testing procedures, I then made appropriate changes in the data collection instruments and procedures. For example, although Kim stated that she liked being given three types of sources, she did not use the graph while writing her essay. Therefore, it was considered necessary to emphasize orally, in addition to the existing task instructions, during the actual data collection that the students should use information from all the sources to support their arguments in their essays. Finally, by conducting a preliminary analysis of the pilot data, I was able to establish that the proposed data analysis methods and procedures could help me answer the stated research questions.

References

- Allen, S. (2004). Task representation of a Japanese L2 writer and its impact on the usage of source text information. *Journal of Asian Pacific Communication*, 14(1), 77–89.
- Benevento, C., & Storch, N. (2011). Investigating writing development in secondary school learners of French. *Assessing Writing*, 16(2), 97–110.
- The College Board (2013). *AP French language and culture 2013 free-response questions*. Retrieved from https://secure-media.collegeboard.org/digitalServices/pdf/ap/apcentral/ap13_frq_french_language.pdf
- Corbeil, G. (2000). Exploring the effects of first- and second-language proficiency on summarizing in French as a second language. *Canadian Journal of Applied Linguistics*, 3, 35–62.
- Cumming, A. (2014). Assessing integrated skills. In A. Kunnan (Ed.), *The companion to language assessment* (pp. 216–229). Malden, MA: Wiley-Blackwell.
- Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing*, 10(1), 5–43.
- Cumming, A., Rebuffot, J., & Ledwell, M. (1989). Reading and summarizing challenging texts in first and second languages. *Reading and Writing*, 1(3), 201–219.
- Flower, L. (1994). *The construction of negotiated meaning: A social cognitive theory of writing*. Carbondale, IL: Southern Illinois University Press.
- Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, 32(4), 365–387.
- Gebriel, A., & Plakans, L. (2009). Investigating source use, discourse features, and process in integrated writing tests. *Spain Fellow Working Papers in Second or Foreign Language Assessment*, 7, 47–84.
- Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1–27). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hirvela, A., & Du, Q. (2013). “Why am I paraphrasing?”: Undergraduate ESL writers’ engagement with source-based academic writing and reading. *Journal of English for Academic Purposes*, 12(2), 87–98.
- Knoch, U., & Sitajalabhorn, W. (2013). A closer look at integrated writing tasks: Towards a more focused definition for assessment purposes. *Assessing Writing*, 18(4), 300–308.
- Li, Y., & Casanave, C. (2012). Two first-year students’ strategies for writing from sources: Patchwriting or plagiarism? *Journal of Second Language Writing*, 21(2), 165–180.
- McCulloch, S. (2013). Investigating the reading-to-write processes and source use of L2 postgraduate students in real-life academic tasks: An exploratory study. *Journal of English for Academic Purposes*, 12(2), 136–147.

- Plakans, L. (2007). *Second language writing and reading-to-write assessment tasks: A process study*. Unpublished Doctoral Dissertation. Iowa City, IA: The University of Iowa.
- Plakans, L. (2015). Integrated second language writing assessment: Why? What? How? *Language and Linguistics Compass*, 9(4), 159–167.
- Plakans, L., & Gebril, A. (2013). Using multiple texts in an integrated writing assessment: Source text use as a predictor of score. *Journal of Second Language Writing*, 22(3), 217–230.
- Rivard, L. P. (2001). Summary writing: A multi-grade study of French-immersion and Francophone secondary students. *Language, Culture, and Curriculum*, 14(2), 171–186.
- Ruiz-Funes, M. (1999). The process of reading-to-write used by a skilled Spanish-as-a-foreign-language student: A case study. *Foreign Language Annals*, 32(1), 45–58.
- Shi, L. (2004). Textual borrowing in second-language writing. *Written Communication*, 21(2), 171–200.
- Spivey, N. N. (1984). *Discourse synthesis: Constructing texts in reading and writing*. Outstanding Dissertation Monograph. Newark, DE: International Reading Association.
- Weigle, S. C. (2004). Integrating reading and writing in a competency test for non-native speakers of English. *Assessing Writing*, 9(1), 27–55.
- Weigle, S. C., & Parker, K. (2012). Source text borrowing in an integrated reading/writing assessment. *Journal of Second Language Writing*, 21(2), 118–133.
- Yang, L., & Shi, L. (2003). Exploring six MBA students' summary writing by introspection. *Journal of English for Academic Purposes*, 2(3), 165–192.
- Yu, G. (2013). From integrative to integrated language assessment: Are we there yet? *Language Assessment Quarterly*, 10(1), 110–114.

A VOT Study on the Acquisition of English Stop Contrasts by Marathi Speakers: The Mutual Influence of L1 and L2

Ari Natarina, University of Iowa

The purpose of this study is to investigate how a person's first language (L1) influences their second language (L2) pronunciation and whether L2 has any effect on L1 pronunciation. The Voice Onset Time (VOT) of English and Marathi utterance-initial stops produced by Marathi speakers who resided in the U.S. were measured. The results of the study are compared to the predictions of the Speech Learning Model of Flege (1992). The results of this study confirm the mutual influence of L2 and L1 in the process of second language acquisition.

Introduction

Plosives in languages are often distinguished by Voice Onset Time (VOT), which is “the timing relation between voice onset and the release of occlusion” (Lisker & Abramson, 1964, p. 387). Voiceless aspirated stops have a longer interval between the stop release and the voice onset (referred to as long-lag VOT) than voiceless unaspirated stops (referred to as short-lag VOT). There are also some utterance-initial stops that have glottal buzz preceding the stop release, which is often referred to as prevoicing or negative VOT.

English has a two-way contrast in stops in utterance-initial position that is often claimed to be between stops that are aspirated with long-lag VOT and stops that are (voiceless) unaspirated with short-lag VOT (Beckman, Jessen, & Ringen, 2013; Iverson & Salmon, 1995). Unlike English, Marathi is usually assumed to be like Hindi which has a four-way contrast in stops (Berkson, 2013; Lisker & Abramson, 1964). Marathi stops can be categorized as voiceless unaspirated with short-lag VOT, voiceless aspirated with long-lag VOT, voiced unaspirated with negative VOT, and voiced aspirated or breathy voice with prevoicing and long-lag VOT (Beckman et al., 2013).

The goal of the study reported in this paper is to investigate how English stops are realized by Marathi speakers and how English influences their production of Marathi stops. Previous studies have looked at how stop contrasts in one language influence the acquisition of stop contrasts in another language. Through the comparison of the VOT of utterance-initial stops

in a person's first language (L1) and their second language (L2), several studies (Caramazza, Yeni-Komshian, Zurif, & Carbone, 1973; Flege & Eefting, 1987a, 1987b; Magloire & Green, 1999) illustrated the interference of L1 on L2.

Caramazza et al. (1973), for example, investigated the perception and the production of stops by Canadian French-English bilinguals in both languages. Note that one language (i.e., French) has prevoicing and no aspiration, whereas the other language (i.e., English) has aspiration and no prevoicing. They found that bilingual speakers have distinct categories for English and French fortis stops.¹ However, the distribution of lenis stops (in English and French) was similar to the VOT distribution of monolingual French speakers: English lenis stops were produced with prevoicing instead of with short-lag VOT. This result illustrates the interference of L1 (French) on L2 (English).

Furthermore, Flege & Eefting (1987a) studied the perception and production of English stops by Dutch speakers and Dutch stops by English speakers. This study showed that more exposure to English resulted in greater VOT differences between English and Dutch alveolar fortis stops. In addition, Flege & Eefting (1987b) analyzed the production of lenis and fortis stops in English and Spanish by bilingual Spanish-English speakers in Puerto Rico. The result of their study showed that the bilingual speakers produced English fortis stops with longer VOT values than their Spanish fortis stops. However, the VOT values of these bilinguals were shorter than the VOT values of native English speakers' fortis stops in the control group. Magloire & Green (1999), in contrast, reported that early Spanish-English bilinguals in their study produced similar VOT values with the English and Spanish monolingual subjects when producing utterance-initial stops in their respective languages; thus, bilinguals have separate representations of English and Spanish stops.

Apart from being influenced by L1 sounds, L2 has also been shown to have an effect on the production of L1. Sancier & Fowler (1997) compared the speech production of a Brazilian speaker after several months-long stays in the U.S. and Brazil. They reported that the VOT values of Brazilian Portuguese fortis stops were five to six milliseconds longer after the speaker's stay in

¹ In a two-way stop contrast language, lenis refers to the less articulated of the two corresponding consonants (typically a voiced consonant), whereas fortis refers to the one that is strongly articulated (typically a voiceless consonant).

the U.S. for several months compared to the VOT values measured after the speaker's visit to Brazil.

Chang (2012) conducted a 5-week longitudinal study on the influence of L2 on L1 in the acquisition of Korean by novice learners whose L1 was English. In his experiments, the participants' pronunciation of Korean and English stops was recorded each week. In a two-week period, the learners' VOT for Korean voiceless aspirated stops reached native-like VOT levels. In contrast, the participants' VOT and fundamental frequency (f0) onset of English fortis stops increased significantly, showing that their English fortis stops were influenced by Korean voiceless aspirated stops, which have longer VOT than English voiceless aspirated stops.

There are three main points that can be derived from the results of these studies. First, L1 has an influence on the acquisition of L2 phonology. Second, L2 has an influence on the production of L1. Third, bilinguals or L2 learners have problems in separating stop categories in their L1 and L2. However, note that these experiments considered the effect of L1 on the acquisition of L2 by comparing VOT production in two languages which have different two-way stop contrasts. One language has prevoicing and no aspiration (e.g., French, Dutch, Spanish), whereas the other has aspiration and no prevoicing (e.g., English). What if the L1 has more stop contrasts than the L2? Can these learners produce accurate pronunciation of L2 stops since they do not have to learn new sounds? We might wonder how Marathi speakers pronounce English stops in word- or utterance-initial position given that Marathi has been claimed to have both types of stops that English has.

By comparing the VOT of utterance-initial Marathi stops and English stops, we can determine whether the results of this study are consistent with the predictions of Flege's Speech Learning Model (SLM), a model that predicts how learners learn or fail to learn L2 sounds (Flege, 1992, 2005). The Speech Learning Model is described in Section 2. Section 3 of this paper describes the participants of this study, the stimuli, and the analysis procedure. The results of this study, the discussion, and the conclusions can be found in Section 4, Section 5, and Section 6, respectively.

Background

Flege (1992, 2005) proposed the Speech Learning Model (SLM) to explain results of earlier studies of the influence of L1 and L2 on each other. There are two main assumptions on which the SLM theory is based. The first assumption is that “the capacities underlying successful L1 speech acquisition remain intact across the life span” (Flege, 2003, p. 328). Flege hypothesized that all L2 learners can eventually perceive the difference between L1 and L2 speech sounds, as long as they receive sufficient input from native speakers for some period of time. It also means that L2 learners can sort these differences into separate categories and produce an output that has the properties of the perceived sounds. The second assumption is that “bilinguals cannot fully separate their L1 and L2 phonetic systems” (Flege, 2003, p. 328). A new L2 sound can be pronounced authentically by all learners who have had sufficient input, while L2 sounds that are similar to L1 sounds cannot be mastered (Flege, Munro, & Skelton, 1992). When an L2 sound is equated (Flege, 1992) with an L1 sound, the L1 and L2 category will assimilate and form a merged L1 and L2 (Flege, 2005).

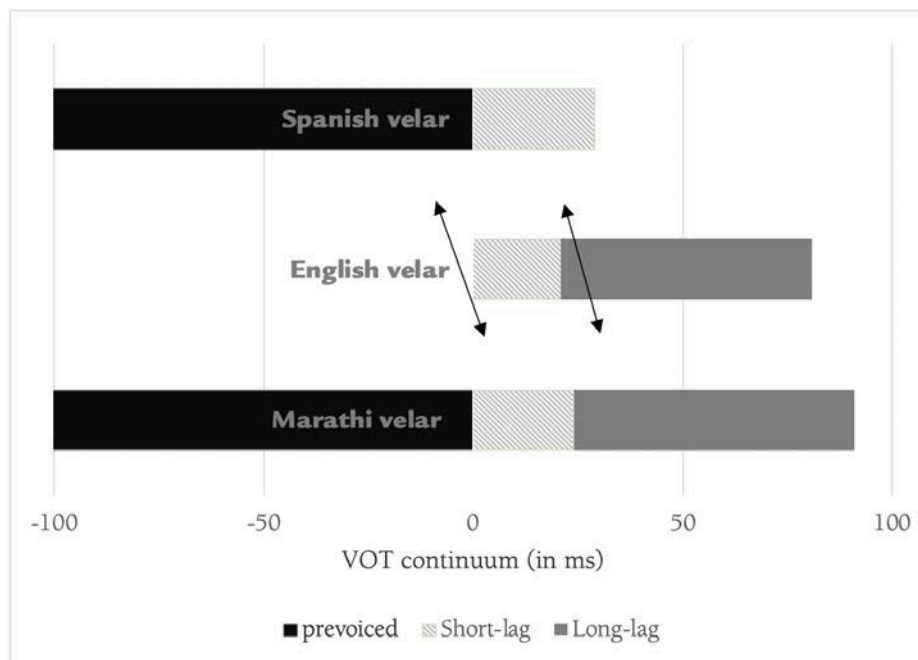
New sounds are the vowels and consonants in the L2 that significantly differ from the sounds available in the L1 inventory, whereas similar L2 sounds are perceived to be equivalent to L1 sounds. For example, French /y/ is considered to be a new sound because there is no equivalent sound in English. On the other hand, French /u/ is considered to be similar to English /u/, although English /u/ has a higher F2 than French /u/. Flege (1987) compared the production of these vowels by English speakers who were advanced learners of French and French monolinguals and found that there was no significant difference in the production of French /y/ by experienced English speakers of French and monolingual French speakers, suggesting that advanced learners of French managed to create a separate category for French /y/. However, the production of French /u/ by English speakers had a higher F2 frequency than the monolingual French speakers, suggesting that they perceived French /u/ to be in the same category as English /u/. These results support the prediction of the SLM.

In the same study, Flege also looked at the production of French /t/ by English speakers (1987). Less experienced English speakers of French produced French /t/ with longer VOT values equal to the VOT values of the English alveolar fortis stop, whereas the more experienced learners of French produced intermediate VOT values shorter than the VOT values of the English

fortis stop and longer than the VOT values of French /t/. Flege argued that this is the effect of equivalence classification, in which learners identify L2 phones (e.g., French or Spanish [t]) as being in the same category as the nearest L1 sounds (i.e., English [t^h]), in the same way as children “identify phones produced by different talkers, or in different phonetic context, as belonging to the same category” (1987, pp. 49–50) during their first language acquisition.

Figure 1 below shows the VOT continuum of stops in different languages and how L2 learners map sounds in L2 to sounds in L1 as the effect of equivalent classification. Because Spanish and English have two categories of stops, Spanish speakers learning English would group together the stops with “the greater positive VOT values” (Holliday, 2015, p. 2), i.e., English fortis stops with long-lag VOT and Spanish voiceless stops with short-lag VOT. In contrast, stops with “the lesser positive VOT values” (Holliday, 2015, p. 2), i.e., English lenis stops with short-lag VOT and Spanish voiced stops with negative VOT, would be categorized as one group.

Figure 1. The Range of VOT Values of Velar Stops in Spanish and English (Lisker & Abramson, 1964) Compared to Marathi Velar Stops



The SLM hypothesis of equivalence classification can offer an explanation for why speakers of languages with a voicing contrast (e.g., Spanish and French) do not achieve native-

like production of stops in an aspirating language (e.g., English), and vice versa. Is the SLM prediction also correct for learners whose L1 has fewer or more stop contrasts than the L2?

Relevant to the issue of stop sounds acquisition by learners whose L1 has fewer stop contrasts than their L2 are the studies carried out by Pisoni, Aslin, Percy, and Hennessy (1982) and Holliday (2014, 2015). Pisoni et al. (1982) carried out a perceptual study on the distinction of three-way voicing contrasts by naïve listeners. The purpose of their experiment was to investigate whether native English speakers can perceive the difference between three sounds in the VOT continuum: a negative, a short-lag, and a long-lag VOT. They found that the participants of the study could recognize the differences between the three different sounds easily after they were given a short training in a laboratory.

Holliday (2014) studied Mandarin speakers' perception of Korean stops. Mandarin is an aspirating language like English, and Korean is a language with a three-way stop contrast between short lag (fortis/tense), intermediate lag (lenis/lax), and long lag (aspirated) stops. The Mandarin speakers perceived Korean lenis and aspirated stops to be similar to each other, while fortis stops were perceived to be different from the other two types of stops. Based on SLM's equivalence classification, it is expected that Mandarin speakers should produce Korean stops according to the way they are perceived: short-lag VOT for Korean fortis stops and long-lag for Korean lenis and aspirated stops. However, Holliday (2015) found some Mandarin speakers conformed to SLM's prediction because they produced Korean fortis stops with short-lag VOT, whereas Korean lenis and aspirated stops were produced with long-lag VOT. In contrast, other Mandarin speakers produced Korean fortis and lenis stops with short-lag VOT and Korean aspirated stops with long-lag VOT, deviating from SLM's prediction. Moreover, there were also Mandarin speakers who produced the three phonation types with overlapping VOT.

The results of Holliday's (2015) production study do not seem to support SLM's prediction of equivalence classification for the production of L2 stops by learners whose L1 has fewer stop categories. Can SLM predict the production of L2 stops by learners whose L1 has the L2 stops among all the stop categories in its sound inventory? The study reported in this paper examines the production of English utterance-initial stops by Marathi speakers, whose L1 has more stop contrasts than their L2. Marathi has four stop categories, two of which are like English stops in utterance-initial position with short-lag and long-lag VOT, respectively. Because the

SLM predicts that L2 learners will associate L2 sounds with the most similar sound in L1, we can expect that Marathi speakers would be able to produce English lenis stops with short-lag VOT and English fortis stops with long-lag VOT.

Method

There were two conditions in this study. In condition I, the participants read a list of Marathi words. The VOT of utterance-initial stops in Marathi was analyzed to establish the VOT representations of Marathi stops and to check if there was influence of L2 in the participants' production of Marathi stops. In condition II, the participants read a list of English words. The VOT of utterance-initial stops in English were analyzed to investigate how Marathi speakers realized the English sounds and to check whether there was L1 influence on the L2 production.

Participants

The participants in this study were eight adult native speakers of Marathi, four males and four females. These participants were living in Iowa City at the time of the study. Other than Marathi and English, the participants also speak Hindi, and several participants are familiar with Gujarati, Tamil, and Sanskrit. The data were recorded in an anechoic chamber at the University of Iowa.

Stimuli

The Marathi stimuli consisted of words with utterance-initial stops in three places of articulation: bilabial, dental, and velar. The VOT for retroflex stops in Marathi was not investigated because there are no corresponding retroflex stops in English. The stimuli for English were controlled for the vowel following the utterance-initial stops because vowels were found to have an effect on the VOT. The English words in the stimuli had vowels that also exist in the Marathi vowel inventory in order to eliminate problems that may come up with pronouncing a non-native vowel.

Stimuli for Condition I (Marathi)

Participants read a word list that consisted of 61 fillers and 24 tokens (listed in Table 1).² The words in the list were written in Devanagari script with an English translation next to them,³ and the participants were asked to read the words in the list twice with a pause in between words. A total of 384 tokens were analyzed, but only 352 were included in the analysis; tokens with utterance-initial bilabial voiceless aspirated stops were discarded since they were produced as fricatives.

Table 1. Marathi Tokens (shown in IPA)

Place of Articulation	Voiced	Voiced Aspirated	Voiceless	Voiceless aspirated
Bilabial	bolne	bhəʈne	pustək	p ^h up ^h us
	bara	b ^h at	pac ^h	p ^h ata
Dental	ɖəgəɖʈ	ɖ ^h əg	ʈəra	ʈ ^h uŋki
	ɖər	ɖ ^h ar	ʈəs	ʈ ^h aʈi
Velar	gəwət	g ^h ər	kudel	k ^h up
	gal	g ^h ar	kak ^h	k ^h as

Stimuli for Condition II (English)

The English stimuli consisted of 12 words with utterance-initial stops (as shown in Table 2) and 78 fillers. There were 192 words analyzed for the VOT of the utterance-initial stops. However, only 190 tokens were included in the analysis; two instances of the velar voiced stop in the word *geese* [ki:z] were discarded because they were pronounced as affricates.

² The analyzed tokens in the Marathi word list are not all minimal pairs. Some of the stimuli have the high round vowel [u] following the utterance-initial stops. It has been observed that vowel context can affect VOT in that the VOT of a stop is generally longer when it is followed by a high vowel or a tense vowel than when it precedes non-high vowels or lax vowels (Neary & Rochet, 1994; Port & Rotuno, 1979; Rochet, Nearey, & Munro, 1987). I found that the VOT of Marathi plain voiceless bilabial stops and voiceless aspirated alveolar stops is longer when the stops are followed by high rounded vowels (with an average VOT difference of 11.5 ms). However, the same trend does not occur for the plain voiceless velar stops and the voiceless aspirated velar stops.

³ The English translation was included in the word list as an extra measure in case the participants did not have the ability to write or read Devanagari script.

Table 2. English Tokens (in IPA)

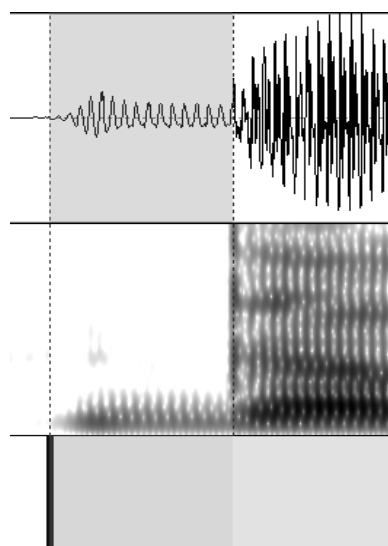
Place of Articulation	Lenis	Fortis
Bilabial	pæt	p ^h æt
	pi:p	p ^h i:p
Alveolar	tæb	t ^h æb
	ti:n	t ^h i:n
Velar	kæb	k ^h æb
	ki:z	k ^h i:s

Acoustic Analysis

Praat was used to visualize waveforms and spectrograms for each analyzed token from the recorded speech.

Measuring prevoicing duration. Jessen (1998) defined negative VOT or prevoicing as “the temporal amount of periodic low energy structure during stop closure immediately preceding stop release” (p. 76). The prevoicing of the utterance-initial plain voiced and voiced aspirated stops was measured from the first periodic sinusoidal wave to the beginning of the release burst in the waveform. Figure 2 illustrates how the negative VOT was measured in this study.

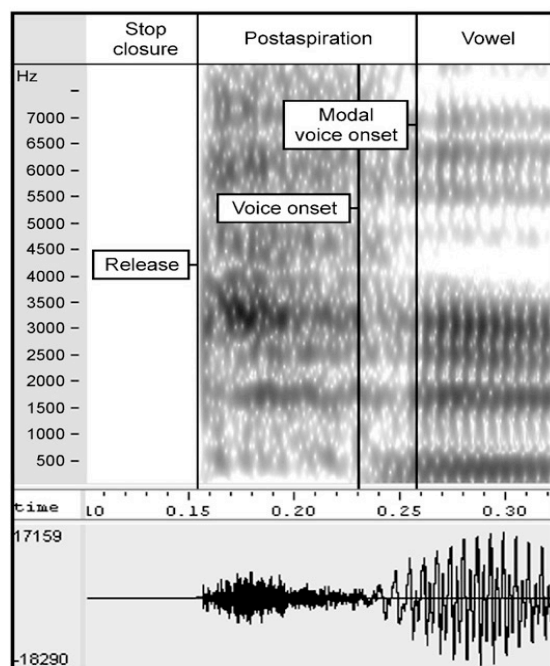
Figure 2. Voicing in Utterance-Initial Position



Measuring Aspiration Duration

Lisker and Abramson (1964) measured VOT from the release of the stop to the onset of glottal vibration or voicing. However, Ladefoged and Maddieson (1996) and Helgason and Ringen (2008) suggested measuring the interval between stop release and the modal voice onset, as shown in Figure 3. I adopted this in measuring the aspiration of utterance-initial Marathi and English stops in this study. Using the spectrogram, I measured the aspiration from the release burst to the first clear formant in the spectrogram, marking formant onset at the nearest zero crossing in the waveform.

Figure 3. An Example of How Aspiration was Measured by Helgason and Ringen (2008, p. 610)

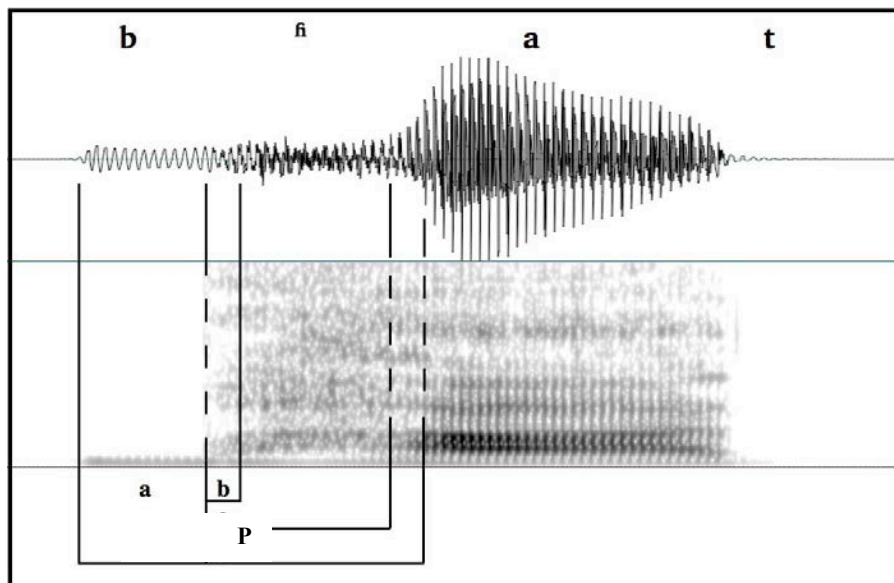


Measuring Breathy Voice Duration

The end of breathy voice or the aspiration in voiced aspirated stops is less clear than the end of aspiration of voiceless stops. Davis (1994) introduced Noise Offset Time (NOT), which is measured from the release burst to the second formant of the following vowel. Mikuteit and Reetz (2007) suggested After Closure Time (ACT), in which the measurement begins at the release burst and ends at the beginning of periodicity in the waveform. Berkson (2012) used the Pre-vocalic Interval (PVI), which refers to the duration beginning “from the release burst to the jump in intensity which coincides with the appearance of clear formants in the spectrogram of

the following vowel” (p. 27). Figure 4 depicts how the NOT, ACT and PVI are measured in waveform and spectrogram. I used Berkson’s method in measuring the aspiration of voiced aspirated stops in Marathi because it is the combination of how I measured prevoicing (Jessen, 1998) and aspiration (Helgason & Ringen, 2008; Ladefoged & Maddieson, 1996).

Figure 4. The Measurements of NOT, ACT, and PVI According to Berkson (2012, p. 42)



Results⁴

The VOT of Marathi Stops

The VOT of Marathi stops produced by the participants in this study (Table 3) were similar to those in Lisker and Abramson (1964; see Table 4), in that the plain voiced stops were prevoiced, the plain voiceless stops had short-lag VOT values, and the voiceless aspirated stops had long-lag VOT values. However, the voiced aspirated stops in this study were measured for prevoicing and aspiration (breathy voice), unlike those in Lisker and Abramson’s study, which were only measured for the negative VOT. In addition, Lisker and Abramson found that the bilabial voiceless aspirated stops had long-lag VOT, but in this study, for the tokens where utterance-initial bilabial voiceless aspirated stops were expected, fricatives were produced by all speakers (except for one speaker’s production of the word *phata*).

⁴ The author thanks Dr. Jun Yin for assistance with statistics in this study.

Table 3. Marathi VOT in this Study (in milliseconds)

Place of Articulation	Plain voiced	Plain voiceless	Voiceless aspirated	Voiced aspirated	
	Neg. VOT	Aspiration	Aspiration	Neg. VOT	Aspiration
Bilabial	-84.89	17.60	-	-70.60	55.31
Dental	-94.97	15.03	59.83	-69.02	62.54
Velar	-80.92	30.30	85.38	-69.24	78.84

Table 4. Marathi VOT (in ms) in Lisker and Abramson (1964)

Place of Articulation	Plain voiced	Plain voiceless	Voiceless aspirated	Voiced aspirated
Bilabial	-117	11	76	-95
Dental	-111	0	63	-87
Velar	-116	24	87	-89

The data in Table 3 shows that Marathi has four stop categories with different VOT ranges. To establish that the four stop types in Marathi are separate stop categories, a two-way ANOVA was used to test the effect of voicing condition (voiced, voiceless) and aspiration condition (aspirated, unaspirated). The result shows that there was a significant main effect of voicing on the VOT values ($F(1,7) = 364.02, p < 0.001$), which means that the VOT values for voiced and voiceless stops were significantly different. Aspiration was also found to have a significant effect on VOT values ($F(1,7) = 551.37, p < 0.001$). This effect tells us that the VOT values were different for aspirated and unaspirated stops. I also found a significant interaction effect between voicing and aspiration on the VOT values of Marathi stops ($F(1,7) = 29.56, p = 0.001$). This statistic shows that the voiced unaspirated stops had different negative VOT values from the voiced aspirated stops, and the lag VOT values of voiceless unaspirated stops were also significantly different compared to the voiceless aspirated stops.

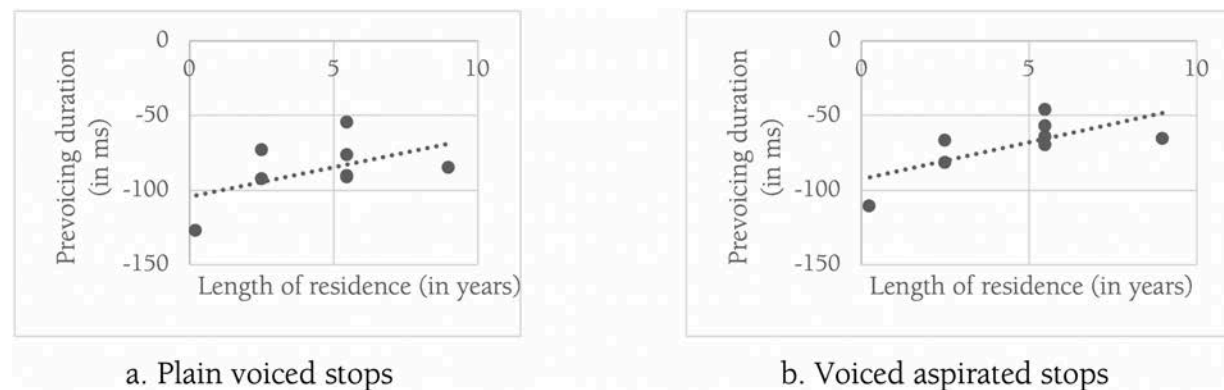
To check whether aspiration in Marathi has different qualities depending on stop types, (i.e., voiceless, voiceless aspirated, and voiced aspirated), a one-way ANOVA was run to compare aspiration in these three stop types. The result shows that the aspiration durations for the three stop types were significantly different ($F(2,14) = 32.79, p < 0.001$). A paired samples t-test was used to compare the means of the aspiration duration for each stop type. The aspiration duration mean for the voiceless unaspirated stops was found to be significantly different from the

aspiration duration mean for the voiceless aspirated stops ($t(7) = -14.40, p < 0.001$). A significant difference was also found when the means of the aspiration duration for voiceless unaspirated and voiced aspirated were compared ($t(7) = -5.00, p = 0.002$). However, the difference between the means of the aspiration duration for the voiceless aspirated and voiced aspirated was found to be insignificant ($t(7) = 0.98, p = 0.35$). This result indicates that the breathy voiced stops in Marathi have true aspiration.

The Influence of English on Prevoicing

To check whether English has an influence on the participants' prevoicing in this study, we can compare the negative VOT values of the speaker who had the shortest length of stay in the U.S. with the negative VOT values of other speakers with longer length of residence (LOR). Since there is no voicing contrast in utterance-initial English stops, we can predict that the prevoicing duration of Marathi speakers who have lived in the U.S. for a longer period of time will be shorter if English has an influence on the negative VOT in participants' Marathi. From Figure 5, we can see that English has apparently influenced the amount of prevoicing in plain voiced and voiced aspirated stops. The trend lines show that the longer Marathi speakers reside in the U.S., the shorter their prevoicing is.

Figure 5. The Mean Prevoicing Duration of Marathi Plain Voiced (a) and Voiced Aspirated (b) Across Groups Based on Length of Residence (LOR)



Aspiration

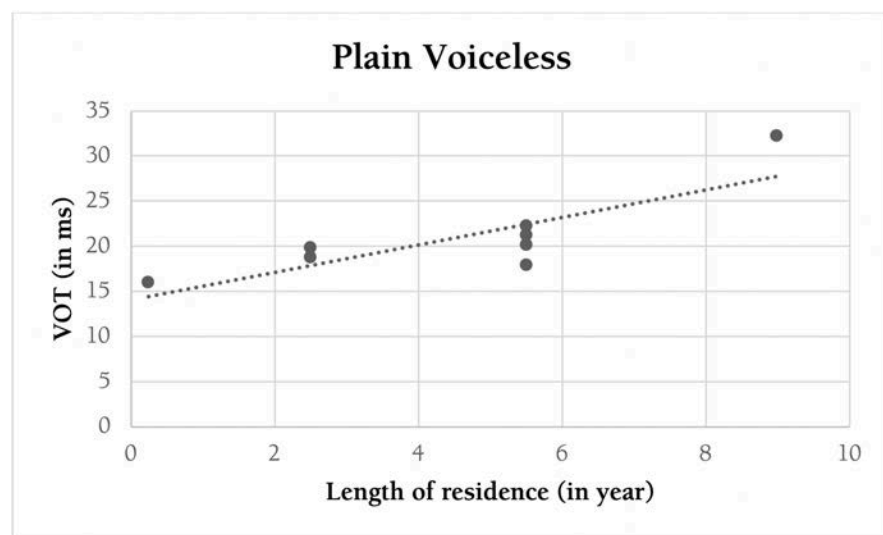
To test whether there is a correlation between how long participants lived in the U.S. with how good their pronunciation of English stops was, I used a simple linear regression model. In the simple linear regression model, the LOR was used as the variable in predicting the VOT

outcomes. For the plain voiced data, no significant regression equation was found ($F(1,6) = 1.72$, $p = 0.23$), with an R^2 value of 0.223, which means that the correlation between LOR and the VOT values of plain voiced stops is not perfectly linear because only 22% of the variance in prevoicing duration can be predicted by LOR. The regression coefficient was 3.64. This indicates that, for each year that the length of residence increases, the prevoicing VOT will shorten by 3.64 ms. However, prevoicing duration of the utterance-initial voiced aspirated can be predicted by the LOR as a significant result was found ($F(1,6) = 7.79$, $p = 0.03$). The R^2 is equal to 0.56. This means that there is not a perfect linear correlation between LOR and VOT. The regression coefficient shows that the participants' negative VOT for voiced aspirated stops will increase by 5.46 ms for each additional year of residence. This result suggests L2 English influences the production of Marathi utterance-initial stops.

The Influence of English on Aspiration

Since both English and Marathi have an aspiration contrast in utterance-initial position, we can expect that there will be no effect of L2 on the L1 production. However, Figure 6 shows that the longer Marathi speakers have lived in the U.S., the longer their VOTs are for Marathi plain voiceless stops.

Figure 6. The Mean VOT Values of Marathi Plain Aspirated Across Groups Based on Length of Residence (LOR)

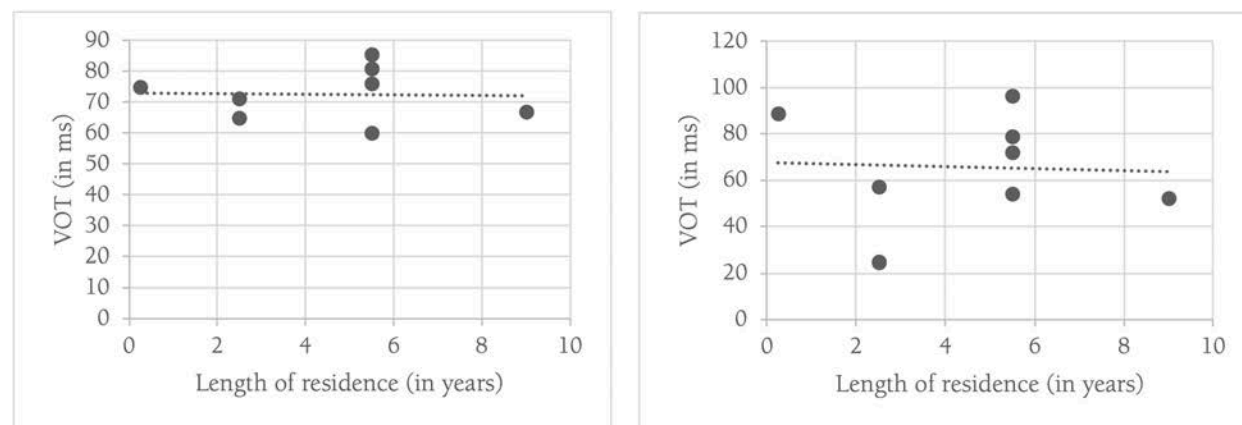


To check if there is a significant LOR effect on the VOT of Marathi plain voiceless stops in this study, a simple linear regression test was carried out. A significant correlation between

LOR and short-lag VOT ($F(1,6) = 12.97, p = 0.01$) was found. The R^2 value of 0.68 shows that LOR and the VOT values of plain voiceless stops have a perfect linear correlation. The regression coefficient value was 1.52, which means that for each additional year of residence, the prevoicing VOT will increase by 1.52 ms. This result portrays the influence of English on the production of Marathi plain voiceless stops, which typically have short-lag VOT values.

However, when we look at the VOT distribution for the voiceless aspirated stops in this study (Figure 7a), we can see that there is no trend across LOR groups because the speakers in the 2.5 LOR group have the shortest average of aspiration duration, while the 5.5 LOR group has the longest average of aspiration VOT values. This means that there is no influence of L2 on Marathi voiceless aspirated stops. Likewise, since the aspiration of Marathi voiceless aspirated stops is not influenced by LOR, we may expect that this variable would not influence the aspiration of Marathi voiced aspirated stops. Figure 7b shows that this is true: There is no effect of LOR on the aspiration of Marathi voiced aspirated stops.

Figure 7. The Mean Aspiration Duration of Marathi Voiceless Aspirated (a) and Voiced Aspirated (b) Stops Across Groups Based on Length of Residence (LOR)



a. Voiceless aspirated

b. Voiced aspirated

A simple linear regression model with LOR as the predictor for the positive VOT outcomes was run. For the voiceless aspirated data, the regression equation was insignificant ($F(1,6) = 0.006, p = 0.93$). Furthermore, there was no significant correlation between LOR and the aspiration duration of voiced aspirated stops. The results ($F(1,6) = 0.01, p = 0.90$) were consistent with the findings from the voiceless aspirated data. These results indicate that, for aspirated stops, there is no L2 influence over L1, particularly in the long-lag category.

The VOT of Marathi Speakers' English Stops

Marathi speakers in this study pronounced 97.9% of the English lenis stops with prevoicing. For English fortis stops, 99% were produced with positive VOT values. The mean VOT of English stops produced by the participants in this study is shown in Table 5. If we compare the results in Table 5 with Lisker and Abramson's (1964) findings for English stop VOT values for English native speakers in Table 6, we see that the results are not as expected.⁵

Table 5. The Mean VOT of English Stops by Marathi Speakers (in ms) in this Study

	Lenis (neg. VOT)	Fortis (pos. VOT)
Bilabial	-104.89	36.17
Alveolar	-103.98	41.28
Velar	-106.1	62.32

Table 6. The Mean VOT of English Stops (in ms) by English speakers in Lisker and Abramson (1964, p. 394)

	Lenis		Fortis (pos. VOT)
	(neg. VOT)	(pos. VOT)	
Bilabial	-101	1	58
Alveolar	-102	5	70
Velar	-88	21	80

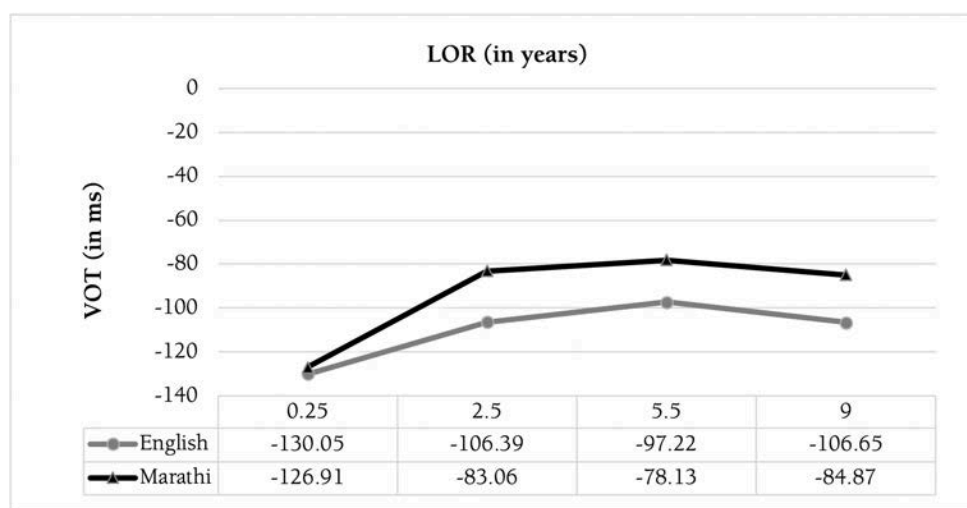
Marathi speakers had negative VOT values for their English lenis stops, and the English fortis stops produced by Marathi speakers have shorter positive VOT values than the English native speakers. These Marathi speakers also have shorter positive VOT values on fortis stops than for the same category in their native Marathi. The influence of L1 on the production of L2 stops in each stop category is analyzed in the following subsections.

⁵ Lisker and Abramson (1964) categorized the VOT of English lenis stops as short-lag VOT and negative VOT because one of their participants prevoiced 95% of the English lenis stops, whereas the other three produced English lenis stops with short-lag VOT values. Only the average from the short-lag VOT category is included in Table 6.

English Lenis Stops

The participants in this study employed prevoicing in their production of utterance-initial English lenis stops, as illustrated in Figure 8. The fact that Marathi speakers had robust prevoicing in pronouncing English lenis stops is apparently evidence for the influence of L1 on L2 acquisition because Midwestern English speakers usually do not have prevoicing utterance-initial lenis stops. This finding is in accordance with the findings of earlier studies of the acquisition of English by speakers of languages with prevoicing like French (Caramazza et al., 1973) and Spanish (Flege & Eefting, 1987b).

Figure 8. The Comparison of the Mean Negative VOT for Marathi Plain Voiced Stops and English Lenis Stops Across LOR Group



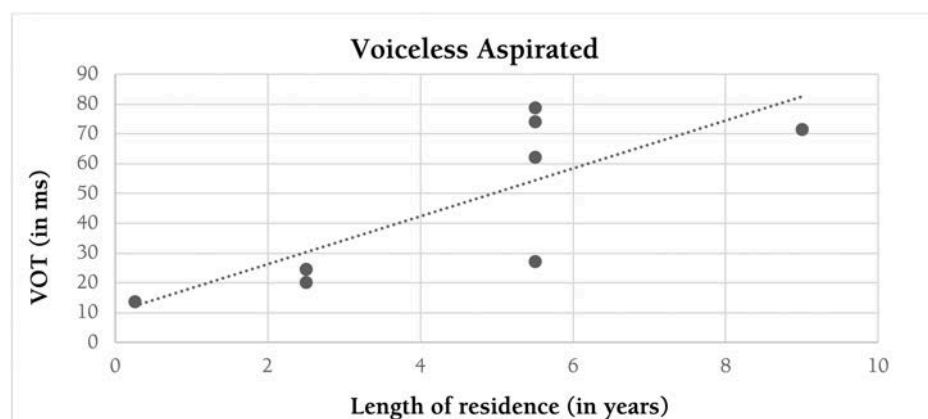
Using a simple linear regression model, the VOT outcomes are predicted by the LOR as a single continuous predictor. For the Marathi speakers' production of utterance-initial English lenis stops, an insignificant regression equation was found ($F(1,6) = 2.42, p = 0.17$). The correlation between the negative VOT of Marathi plain voiced stops and English lenis stops in this study is estimated using Pearson's correlation test as 0.54 (95% confidence interval [0.17, 0.77]), which means that the VOT values for these two stop categories are positively correlated. If a Marathi speaker has long prevoicing for Marathi stops, s/he will have the tendency to have long prevoicing for the English lenis stops. Furthermore, a paired t-test was carried out to test whether the mean of Marathi plain voiced stops prevoicing duration and the mean of English lenis stops prevoicing duration in the three places of articulation were the same. The prevoicing durations for the Marathi and English stops in the three places of articulation were found to be

significantly different ($t(23) = -4.42, p = 0.00019$). Even though they are positively correlated, the Marathi plain voiced stops have different prevoicing lengths from the English lenis stops.

English Fortis Stops

Figure 9 shows that the VOT values of utterance-initial English fortis stops produced by Marathi speakers increased as LOR increased. This result indicates the influence of L1 on L2 because Marathi speakers who had less English input from native English speakers (shorter LOR) pronounced utterance-initial English fortis stops with short-lag VOT, whereas those who had lived in the U.S. longer and had received more accurate input pronounced the utterance-initial English fortis stops with long-lag VOT values.

Figure 9. The Mean VOT of English Fortis Stops by Marathi Speakers Across LOR Groups

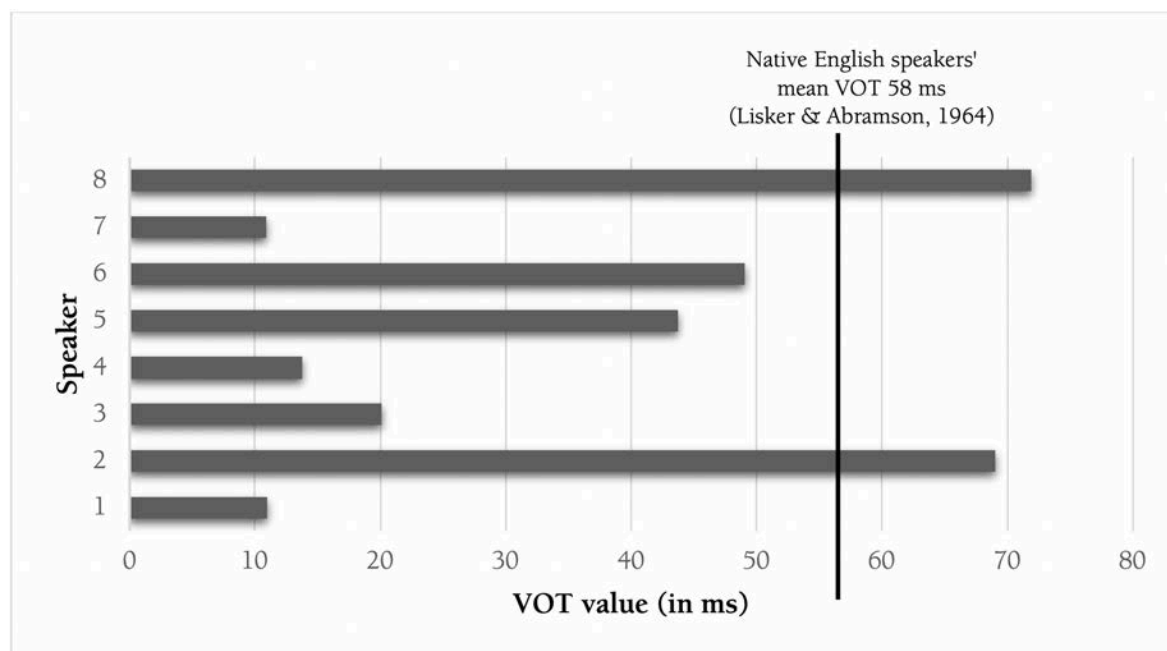


The correlation between the length of VOT and LOR for all the speakers was found to be significant through a simple linear regression test ($F(1,6) = 12.59, p = 0.01$). The R^2 value was 0.67, suggesting that 67% of the variance in VOT can be explained by the LOR. In other words, the VOT values of English fortis stops by Marathi speakers is strongly predicted by how long they have lived in the U.S. The regression coefficient of the residence effect shows that participants' VOT can be predicted to increase by 8.2 ms for each year they stay longer in the U.S. This outcome supports the SLM hypothesis, which predicts that the more a learner gets input from native speakers, the more likely it is that s/he will be able to produce a native-like L2 sound.

Evidence bearing on the SLM prediction that acquisition of a new L2 sound is less challenging than acquisition of a sound similar to an L1 sound is provided by the pronunciation

of English bilabial fortis stops. As noted earlier, Marathi speakers do not produce bilabial voiceless aspirated stops, rather they produce fricatives (except for speaker 7's production of the token phata); therefore, the English bilabial fortis stop can be considered a new sound for Marathi speakers. SLM predicts that speakers should be able to acquire this new sound; hence, they should all have long-lag VOT for English bilabial fortis stops. Figure 10 shows that not all speakers acquired this new sound. Only speakers 2, 5, 6, and 8 had long-lag VOT for the English bilabial fortis stops, whereas the other speakers produced stops with short-lag VOT.

Figure 10. The Comparison of English Bilabial Fortis VOT (in ms) by Marathi-English Speakers

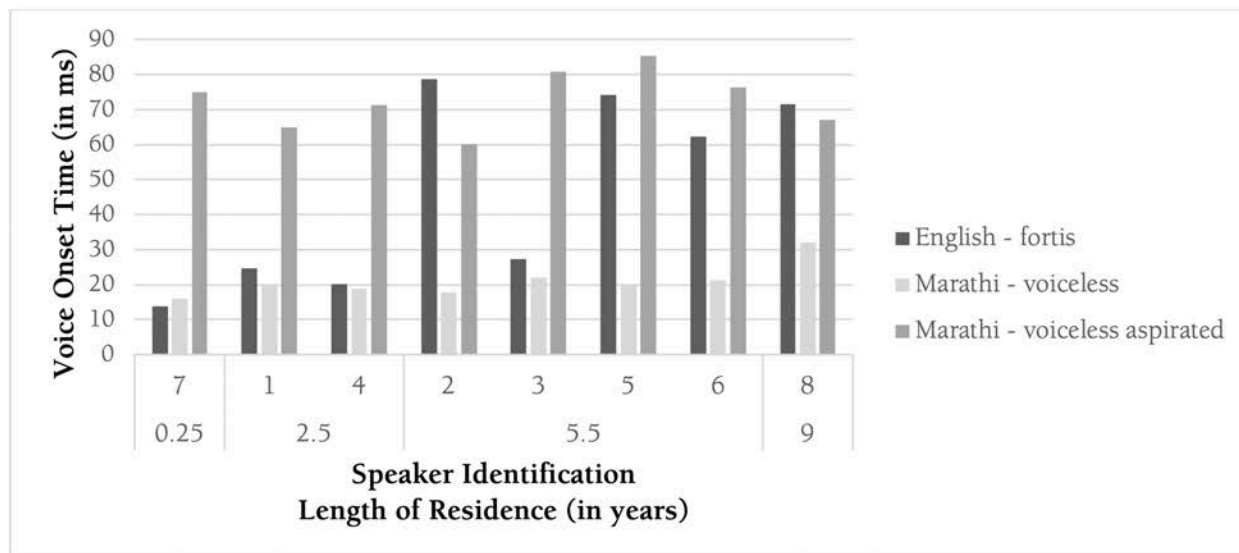


Consequently, we should also look at the mean VOT of English fortis stops for individual speakers to see how each speaker realizes their English fortis stops. The comparison of the mean VOT of English fortis, Marathi plain voiceless, and Marathi voiceless aspirated stops produced by the participants is given in Figure 11 below.⁶ This comparison shows that a Marathi speaker either produces all English fortis stops like Marathi voiceless aspirated stops or like Marathi voiceless unaspirated stops. The production of English fortis stops by speakers 1, 3, 4, and 7 was

⁶ The mean VOT for Marathi voiceless aspirated stops in Figure 11 consisted of only dental and velar stops.

similar to their L1 plain voiceless stops. In contrast, speakers 2, 5, 6, and 8 produced their English fortis stops similar to voiceless aspirated stops in Marathi.

Figure 11. The Comparison of Mean VOT Values (in ms) Between English Fortis Stops, Marathi Plain Voiceless, and Marathi Voiceless Aspirated Stops for Individual Speakers



Accordingly, a one sample t-test was used to determine if Marathi speakers pronounced English fortis stops like English native speakers by comparing each Marathi speaker's mean VOT production for English fortis stops with the mean VOT of the English native speakers listed in Lisker and Abramson (1964). In other words, this test was carried out to clarify which participants produced English fortis stops appropriately (with the long-lag VOT) and inappropriately (with the short-lag VOT). Similar to the illustration in Figure 11, the results of the one sample t-test confirmed that English fortis stops produced by Marathi speakers were different from the native English speakers' fortis stops in the case of the first participant ($t(11) = -10.2, p < 0.0001$), the third participant ($t(10) = -4.7, p = 0.0008$), the fourth participant ($t(11) = -10.2, p < 0.0001$), and the seventh participant ($t(11) = -26.8, p < 0.0001$). In contrast, the second participant ($t(11) = 1.3, p = 0.21$), the fifth participant ($t(11) = 0.6, p = 0.56$), the sixth participant ($t(11) = -1.4, p = 0.18$), and the eighth participant ($t(11) = 0.5, p = 0.63$) produced English fortis stops with native-like VOT values.

Discussion

The Influence of Marathi on English Stops

The SLM predicts that Marathi speakers should be able to pronounce English lenis and fortis stops in a native-like manner, with short-lag and long-lag VOT values, because these L2 sounds exist in their L1 sound inventory. However, the findings in the current study do not support this prediction. Marathi speakers produced two different contrasts for English: Four participants pronounced English utterance-initial stops with a negative VOT—short-lag VOT contrast, while the other four produced English stops with a negative VOT—long-lag VOT contrast.

Why did the Marathi speakers in this study have negative VOT instead of short-lag VOT values for their English lenis stops? Prevoicing seems to be a consistent phenomenon in the acquisition of English sounds by speakers of other languages. French-English bilinguals (Caramazza et al., 1973) and Spanish-English bilinguals (Flege & Eefting, 1987b) employed prevoicing for their English lenis stops. These studies, however, investigated the acquisition of English stop sounds, which have aspiration as the contrast, by speakers whose L1 has two types of stops with a voicing contrast. Therefore, the SLM hypothesis of equivalence classification can explain the results of these previous studies. Speakers of Spanish and French equate the short-lag VOT English lenis stops with the most similar sound in L1, which is the prevoiced Spanish/French lenis stops.

The fact that Marathi speakers had prevoicing in their pronunciation of English lenis stops cannot be explained by the SLM hypothesis of equivalence classification. If equivalence classification holds, Marathi speakers should be able to produce English lenis stops with short-lag VOT values because the closest equivalent to English lenis stops is Marathi plain voiceless stops. One possible explanation for why speakers of Marathi (and other languages with a voicing contrast) prevoiced their English lenis stops is that English lenis stops undergo passive voicing; thus, they are sometimes voiced in an intervocalic position (Beckman et al., 2013).⁷ To support their hypothesis of passive voicing in aspirating languages, Beckman et al. compared the percentage of intervocalic lenis stops in languages with a voicing contrast (i.e., Russian) to

⁷ Beckman et al. (2013) described passive voicing as a phonetic process where voicing happens because of the voiced context.

intervocalic lenis stops in languages with an aspiration contrast (i.e., German). They found that 62.5% of German lenis stops in intervocalic contexts were voiced for more than 90% of the closure, while more than 97% Russian lenis stops were fully voiced in intervocalic positions. This passive voicing process does not occur in Marathi because, in addition to an aspiration contrast between stop categories, it also has a voicing contrast like Russian. Therefore, instead of alternating between negative and short-lag VOT values for English lenis stops in different contexts, Marathi speakers might categorize English lenis stops in word-initial and intervocalic position as plain voiced stops. In other words, the intervocalic variants of the lenis stops in English are the basis for equivalence classification, which leads Marathi speakers to perceive utterance-initial English lenis stops as prevoiced.

The transliteration of Devanagari scripts into Roman script in daily written communication may be another explanation for why participants in this study produced English lenis stops with prevoicing. Holliday (2015) mentioned the influence of orthography as one of the reasons for a variety of output in the production of Korean stops by Mandarin speakers. Marathi is written in Devanagari script, which has more consonant letters than the Roman script. Based on the National Library of Kolkata and ITRANS of IIT Madras India (Dhore, Dixit, & Dhore, 2012), Marathi plain voiced stops have the same Roman script representation as English lenis stops. This might influence Marathi speakers' perceptual classification of English sounds to match the symbols *b*, *d*, and *g* in English words with the plain voiced sounds in Marathi. Hence, they produced utterance-initial English lenis stops with prevoicing.

The second question that we need to answer is: why did half of the population in this study equate English fortis stops with Marathi plain voiceless stops, whereas the other half produced English fortis stops similar to Marathi voiceless aspirated stops? The fact that Marathi speakers who have lived in the U.S. longer pronounced their English fortis stops with VOT values similar to Marathi voiceless aspirated stops indicates that they are aware of the aspiration in English fortis stop sounds.

An earlier study by Flege (1987) showed similar results: Native speakers of English who were the least experienced learners of French perceived the French fortis dental stop [t̪] as similar to the English fortis alveolar stop [t^h], thus producing it with long-lag VOT values. In contrast, more experienced French learners perceived a difference between the French

unaspirated stop and the English aspirated stop because they produced the French unaspirated stop with a shorter positive VOT value than the English aspirated stop. The SLM was proposed under the assumption that all learners have the ability to modify their phonetic representation throughout life. Thus, we can expect that the Marathi speakers who produced English fortis stops with short-lag VOT values would be able to pronounce these sounds with long-lag VOT values after they have lived in the U.S. longer. Nonetheless, the SLM does not explain why not all Marathi speakers are immediately able to produce native-like English fortis stops given that Marathi has aspirated stops.

The Influence of English on Marathi Stops

This study presents evidence for the influence of English on the production of Marathi plain voiced and voiced aspirated stops. The results show that the duration of prevoicing is significantly correlated with the speakers' LOR. The prevoicing duration becomes shorter the longer a Marathi speaker lives in the U.S. Because English native speakers do not have prevoicing for lenis stops in utterance-initial position, according to the SLM, Marathi speakers will employ less prevoicing in their Marathi due to the English input they have received.

Flege and Eefting (1987b) investigated the VOT values of utterance-initial English stops and Spanish stops. The results of their study showed that Spanish speakers who learned English earlier in life had shorter prevoicing compared to monolingual Spanish speakers and those who learned English later in life. Although the influence of English was in the direction predicted by the SLM, they found that the group effect was not significant, possibly because their participants received non-native-like English input because they lived in Puerto Rico.

From the results of the current study, English also has an effect on the short-lag VOT stops in Marathi. Marathi speakers who lived in the US longer had longer short-lag VOT compared to those who lived in the U.S. for a shorter time. This result is consistent with the SLM hypothesis that there is a reciprocal influence between L1 and L2. Since English has aspiration to create contrast between stops, it is possible that this contrast influences the production of Marathi plain voiceless stops, resulting in longer positive VOTs for Marathi plain voiceless stops because of the long-lag VOT in English fortis stops. Flege (1987) found similar results in an experiment where experienced English speakers of French had shorter VOT values for their

English alveolar fortis stops than monolingual English speakers and less experienced English speakers of French. In contrast, French participants had longer VOTs for French alveolar fortis stops compared to monolingual French speakers after living in the U.S. for a period of time. These results show that learning an L2 affects learners' production of their L1.

Conclusion

The purpose of this study was two-fold: first, to investigate the acquisition of English, a two-way stop contrast language, by native speakers of Marathi, a language with a four-way stop contrast. The results show that the participants' L1 influences their L2 stop production because English lenis stops (with short-lag VOT) are equated with Marathi plain voiced stops (with prevoicing). In contrast, the production of English fortis stops by Marathi speakers shows inconsistencies: Half of the participants produced English fortis stops with short-lag VOT values, whereas the other half of the participants had long-lag VOT values. Length of residence was found to be a significant factor that influenced Marathi speakers' production of English stops. The second purpose of this study was to investigate the influence of English on Marathi stops. The results show some L2 influence on L1, in that English influences the prevoicing (negative VOT) and short-lag VOT values of Marathi stops.

References

- Beckman, J., Jessen, M., & Ringen, C. (2013). Empirical evidence for laryngeal features: aspirating vs. true voice languages. *Journal of Linguistics*, 49(2), 259–284.
- Berkson, K. H. (2012). Capturing breathy voice: Durational measures of oral stops in Marathi. *Kansas Working Paper in Linguistics*, 33, 27–46.
- Berkson, K. H. (2013). *Phonation types in Marathi: An acoustic investigation* (Unpublished doctoral dissertation). University of Kansas, Lawrence, KS.
- Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *The Journal of Acoustical Society of America*, 54(2), 421–428.
- Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics*, 40(2), 249–268.
- Davis, K. (1994). Stop voicing in Hindi. *Journal of Phonetics*, 22(2), 177–193.
- Dhore, M. L., Dixit, S. K., & Dhore, R. M. (2012). Hindi and Marathi to English NE transliteration tool using phonology and stress analysis. *Proceedings of COLING 2012: Demonstration Papers*, 111–118.
- Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, 15, 47–65.
- Flege, J. E. (1992). Speech learning in a second language. In C. Ferguson, L. Menn, & C. Stoelgammon (Eds.), *Phonological development: Models, research, and application* (pp. 565–604). Timonium, MD: York Press.
- Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In N. O. Schiller, & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 319–355). Berlin, Germany: Mouton de Gruyter.
- Flege, J. E. (2005, April 14–15). *Origins and development of the Speech Learning Model*. Keynote presented at the 1st ASA Workshop on L2 Speech Learning, Simon Fraser University, Vancouver, BC. Retrieved from http://www.jimflege.com/files/SLMvancouver_updated.pdf
- Flege, J. E., & Eefting, W. (1987a). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication*, 6, 182–202.
- Flege, J. E., & Eefting, W. (1987b). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, 15, 67–83.
- Flege, J. E., Munro, M. J., & Skelton, L. (1992). Production of the word-final English /t/ - /d/ contrast by native speakers of English, Mandarin, and Spanish. *Journal of the Acoustical Society of America*, 92, 128–143.
- Helgason, P., & Ringen, C. (2008). Voicing and aspiration in Swedish stops. *Journal of Phonetics*, 36, 607–628.

- Holliday, J. J. (2014). The perceptual assimilation of Korean obstruents by native Mandarin listeners. *Journal of the Acoustical Society of America*, 135(3), 1585–1595.
- Holliday, J. J. (2015). A longitudinal study of the second language acquisition of a three-way stop contrast. *Journal of Phonetics*, 50, 1–14.
- Iverson, G. K., & Salmon, J. (1995). Aspiration and laryngeal representation in Germanic. *Phonology*, 12(3), 369–396.
- Jessen, M. (1998). *Phonetics and phonology of tense and lax obstruents in German*. Amsterdam, Netherlands: John Benjamins.
- Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford, England: Blackwell.
- Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20, 384–422.
- Magloire, J., & Green, K. P. (1999). A cross-language comparison of speaking rate effects on the production of voice onset time in English and Spanish. *Phonetica*, 56, 158–185.
- Mikuteit, S., & Reetz, H. (2007). Caught in the ACT: The timing of aspiration and voicing in Bengali. *Language and Speech*, 50(2), 247–277.
- Neary, T. M., & Rochet, B. L. (1994). Effects of articulation on vowel context on VOT production and perception for French and English stops. *Journal of the International Phonetic Association*, 24(1), 1–18.
- Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, 8(2), 297–314.
- Port, R. F., & Rotunno, R. (1979). Relation between voice-onset time and vowel duration. *Journal of the Acoustic Society of America*, 66(3), 654–662.
- Rochet, B. L., Nearey, T. M., & Munro, M. J. (1987). Effects of voicing, place, and vowel context on VOT for French and English stops. *Journal of the Acoustical Society of America*, 81, S65.
- Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25, 421–436.

An Exploration of L2 Chinese Learners' Perceptions of Interacting with Recap

Runqing Qi, University of Iowa

The purpose of this study is to investigate L2 Chinese learners' perceptions of interacting with Recap (a web-based video prompt online discussion platform) utilizing an online discussion activity that integrated three communicative modes (presentational, interpretive, and interpersonal). Five third-year Chinese college students participated in the study. They first responded to an online technology and language background survey, then used Recap to post video responses, watch others' videos, and leave written comments online. Each participant then individually completed a 15-minute semi-structured interview. The results indicate that all the participants hold a positive attitude toward both the platform (Recap) and the activity. Specifically, the re-record function of Recap allowed learners to practice and optimize their speaking performance by intentionally improving the quality of their speech production through which a sense of achievement can be gained. Additionally, Recap also proved to be beneficial for practicing interpretive listening. By listening to others' videos carefully, participants learned and recalled vocabulary and sentence patterns used by other learners. The interpersonal online discussion through posting videos and textual comments created more opportunities for learners to get to know and encourage each other, which facilitated the formation of a language learning community. Based on participants' general perceptions of the activity, it is suggested to provide individualized feedback for each learner's video, select class-related topics for discussion, and assign the activity as a graded assignment.

Along with the development of computer-assisted language learning (CALL) tools and web-based communication platforms, more and more second language (L2) instructors have started applying technological tools in their teaching practices, and the related research in this field is experiencing a correspondingly prosperous development. Previous studies have reported various benefits that L2 learners have gained through the use of technology. For example, it has been widely proved that the use of technology can improve L2 learners' language skills in reading (e.g., Beauvois, 1994; Gilbert, 2017; Lin, 2016), writing (e.g., Abrams, 2003; Lin & Griffith, 2014), speaking (e.g., Litman, Strik, & Lim, 2018; Swanson & Schlig, 2010) and grammar (e.g., Fazal, Majoka, Khan, & Masood, 2017; Li, 2016; Wang & Smith 2013).

Literature Review

Within this vibrant ongoing CALL research agenda, studies focusing on exploring L2 learners' perceptions, experiences, and attitudes toward the use of technology in L2 learning contexts are receiving increasing attention (e.g., Hsu, Wang, & Comac, 2008; Lee, 2005; Levy, 2015; Yang & Chen, 2007). More and more researchers have begun to realize the significance of investigating this topic: By listening to and analyzing the viewpoints of L2 learners, we can more closely observe and more fully understand their interactive experience with technology. Based on this, L2 instructors can more thoroughly reflect on their pedagogical decisions surrounding the use of technology and make modifications or improvements accordingly to better meet learners' interests and needs.

Among the studies exploring L2 learners' perceptions, it has been widely documented that learners usually hold a positive attitude toward integrating technology in their L2 learning because of the communicative, collaborative, and interactive nature of some of the tools or platforms (e.g., Lee, 2004; Levy & Steel, 2015; Rashed, 2009). Some relatively early literature on this topic (e.g., Lee, 2005; Wang & Coleman, 2009; Yang & Chen, 2006) witnessed a shift of learners' views from regarding technology mainly as an entertaining tool to a beneficial and useful language learning tool; correspondingly, learners become aware of the pedagogical situation that new learning strategies other than the traditional pencil-and-paper learning method should be developed.

It is noteworthy that many studies have reported that the role of the instructor in CALL contexts is considered crucial in forming L2 learners' perceptions. The role of the instructor is critical when giving individualized feedback, coordinating CALL activities, and facilitating the effectiveness and efficiency of the e-learning experience in general. (e.g., Hsu, Wang, & Comac, 2008; Rashed, 2009; Stepp-Greany, 2002; Xu & Peng, 2017)

In particular, among studies analyzing learners' perceptions, investigators have compared synchronous (real-time) technology-assisted communications (such as audio/video-conference and electronic chat) with asynchronous (delayed) technology-assisted communications (such as e-mails and blogs). Although synchronous communication has its irreplaceable merits (e.g., Lee, 2004), some studies (e.g., Ma, 2006; Sotillo, 2000; Warschauer, 1996) have found that learners hold a positive attitude toward asynchronous communication because it creates a relatively

relaxed environment that can relieve foreign language anxiety, especially when they are communicating with L2 native speakers. Learners have additionally reported that due to the delayed nature of asynchronous communication, they have more time to organize their ideas and to improve the complexity and accuracy of their language use. They also receive more opportunities to pay careful attention to and learn vocabulary and structures other participants (either native speakers or L2 learners) use in their expressions. In connection with these previous studies, the present study will focus on web-based asynchronous communication activities.

The involved technology-assisted or web-based language learning tasks in the related literature can be categorized into three types based on different modes of communication (ACTFL, 2012): presentational (speaking/writing; productive) tasks, interpretive (reading/listening; receptive) tasks, and interpersonal (reading and writing/speaking and listening; interactive) tasks. Table 1 lists some examples of activities and technological tools used in different studies given each type of task.

Table 1. A Summary of L2 Learning Tasks and Involved Technological Tools in Learners' Perceptions-Oriented Studies

Task type	Task	Technological tool	Target L2	Study
Presentational	Recording oral assignments	Google's free blogging service: http://www.blogger.com Online audio recording system Evoca: http://www.evoca.com	English	Hsu, Wang, & Comac (2008)
		Wechat (a smartphone social communication application)	Chinese	Xu & Peng (2017)
	Online collaborative writing	PBwiki: e.g., MediaWiki & Google Docs	Spanish	Elola & Oskoz (2010)
Interpretive	Exploring topic-oriented information	WebQuests	English	Fernández (2007)
	Watching pedagogically supported authentic L2 video clips	LoMásTv: http://lomstv.com	Spanish	Pardo-Ballester (2012)
	Using electronic language dictionaries	e.g., WordReference: http://www.wordreference.com (supports a variety of languages) Nciku: www.nciku.com (L2 Chinese)		Levy & Steel (2015)
Interpersonal	Text-based online topic-oriented discussion with native speakers	Virtual Classroom on Blackboard	Spanish	Lee (2004)
		Email	English	Ma (2006)
	Oral communication with native speakers	Video-conferencing	German	Coverdale-lones (2000)

In particular, Lee's (2004) study emphasized the various ways in which discussion around open-ended questions focusing on specific topics can be a powerful way of facilitating L2 learning. First, it promotes meaningful and authentic language use through information exchanging and negotiating. Second, it stimulates learners to participate in the discussion if they are familiar with and interested in the discussed topic. Additionally, focused topics require the use of target or specific vocabulary and structures, so learners will have more opportunities to use and internalize the relevant linguistic constructions they have learned inside or outside of class. Based on these inspiring findings, in this study, an online discussion based on an open-ended question ("What are the differences between Chinese and American campus culture?") was conducted. This topic is related to the content of a reading in Chinese class.

Furthermore, other studies have investigated the holistic e-learning environments created by various combinations the three types of learning tasks using a variety of technological tools or web-based platforms (e.g., Conole, 2008; Rashed, 2009; Stepp-Greany, 2002; Yang & Chen, 2007). Nonetheless, given practical and economic concerns, for both language learners and instructors, it would be more convenient and efficient to manage different types of tasks by using a single technological tool. Nowadays, there are some new web-based platforms that support asynchronous online discussions through both video responses and text-based comments such as Recap (<https://letsrecap.com/>), which may support a combination of presentational, interpretive and interpersonal tasks. Although the integration of different communicative modes is considered important and beneficial in second language learning process, to the researcher's best knowledge, no previous studies have investigated learners' perceptions of interacting with this kind of web-based platform, which incorporates activities of all three communicative modes (e.g., Davin, Troyan, & Hellman, 2011; Nuessel & Marshall, 2008; Tedick & Cammarata, 2006).

An Introduction to Recap

Generally speaking, Recap is an online question-led chat platform designed especially for educational purposes. Teachers can post individual or grouped questions to start a Queue, students can join the Queue through a code shared by the teacher and participate by recording and posting their video responses to the question. After posting the video responses, students can view other participants' videos and type comments to continue the threaded discussion.

Recap facilitates presentation and discussion surrounding a certain topic between students and teachers as well as among students themselves. It also provides teachers with a convenient platform to check students' understanding and give timely feedback (via the posting of comments). Recap is preferable among many similar online platforms because it is free, does not require updates, is accessible through both laptops and mobile devices, supports unlimited queues and participants, and can be integrated with several course-management systems such as Google Classroom, Canvas, and Blackboard.

In terms of modes of communication, holistically, an online discussion activity supported by Recap can be viewed as an interpersonal communication task, in that it promotes interpersonal information sharing and negotiating. Nonetheless, this kind of Recap-supported asynchronous online discussion activity includes several separable steps, such as recording a video response up to five minutes in length (presentational speaking), watching and listening to others' videos (interpretive listening), and posting, reading, and responding to typed comments (interpersonal reading and writing). From a micro view, all three communicative modes are included during the process of completing a Recap-supported online discussion.

Based on the above discussion and introduction, in this study, the following three research questions are proposed:

1. What are Chinese L2 learners' perceptions of using Recap to practice presentational speaking by recording video responses?
2. What are Chinese L2 learners' perceptions of using Recap to practice interpretive listening by viewing other classmates' videos?
3. What are Chinese L2 learners' perceptions of using Recap to do interpersonal online discussions by posting video responses and textual comments?

Method

Participants

Five participants were recruited from a public university in the Midwestern U.S. All the participants were enrolled in a third-year Chinese class during the time of this study. They are all female and native speakers of English.

Table 2. Participants' Information

Participants (in pseudonyms)	Age	Education	Years of studying Chinese
EM	51	Graduate	3
AW	30	Undergraduate	3
SD	21	Undergraduate	3
EC	19	Undergraduate	6
MS	20	Undergraduate	6

Table 2 shows the five participants' age, academic status (when this study was conducted), and how long they had been studying Chinese. The survey indicated that all the participants had had some prior experience of using web-based technologies (such as online dictionaries, Quizlet, Google translator and Google docs, YouTube, etc.) to complete language learning activities; however, the interviews demonstrated that none of the participants had ever used online video prompted discussion platforms such as Recap or before participating in this study.

Procedure

First, participants completed an online survey about their background in L2 learning and technology use. Next, they interacted with Recap by recording and posting a two-minute video to answer the question "In your opinion, what are the differences between Chinese and American campus culture?" As aforementioned, this question is related to a topic that had been the focus of a previous Chinese class. After all the participants had posted their videos, they watched other participants' videos and posted comments on them (further responses to the comments were encouraged). Step-by-step instructions for this Recap interaction activity, which also included some language prompts, were sent to the participants via e-mail (see Appendix A for detailed information). The participants completed the activity in their free time and at their own pace. After all the participants completed the activity, a 15-minute semi-structured interview was scheduled and conducted with each participant individually (see Appendix B for an outline of interview questions).

After each participant had posted her video response, the researcher gave detailed feedback on language form (vocabulary, grammar, and sentence structure), pronunciation (if

necessary), and speech structure after reviewing the video response. All the feedback was posted as public comments; therefore, it was accessible to all the participants.

The interviews were conducted in order to examine the participants' interaction experience with Recap and to understand in detail their perceptions toward this activity in terms of practicing presentational speaking, interpretive listening, and interpersonal reading and writing, as well as general perceptions or potential suggestions toward the whole activity. All the interviews were audio recorded and transcribed, and the transcription was coded. Main categories were developed and organized around four subheadings: presentational speaking, interpretive listening, interpersonal online discussion, and general perceptions of the whole activity.

Results

Presentational Speaking

Generally speaking, all the participants indicated that it was very helpful to practice presentational speaking through recording video responses on Recap. After recording a video, Recap allows participants to replay it and re-record a new one until the participants feel satisfied with their performance. They can then upload and post a finalized version of the video. Accordingly, during the interview, all five participants reported that they had re-recorded their video responses (two to five times) before they finally posted it. Table 3 presents some examples of interview excerpts from the participants, which illustrate why they had re-recorded their videos, as well as more general comments on this presentational speaking task.

Table 3. Interview Excerpts of Participants' Perceptions of the Presentational Speaking Activity

Participants	Excerpts	Themes
EM	(a) ...until I didn't stumble too much in my speaking	Fluency
MS	(b) The first one I was like uh uh uh...	
EC	(c) ...I wanted to redo it so that everything was like nice and smooth.	
EM	(d) ...forced to review the grammar pattern and cemented into your head a little bit more, so it becomes more natural, and you don't have to think about it as much.	
AW	(e) I would go and look up tones.	Accuracy
SD	(f) ...you can play it back obviously and you can hear the mistakes or whatever you want to fix	

Participants	Excerpts	Themes
EM	(g) ...if I left out some big chunks of what I planned to say.	Content organization
AW	(h) Actually really helpful, do it, see it, do it again... replay some part of it, this is what I meant to say, this is how I should have said that.	
SD	(i) I messed up in my thoughts were like not clear.	
EC	(j) To rewrite some parts of my notes, make it clearer for myself	
AW	(k) Because the first time I was done in actually like a minute and so, I said OK, I need to look at some more information and talk a little bit more.	Complexity
	(l) in the classroom...a lot of times we can get away with partial answers, like a noun...this kind of broken Chinese going on. (m) (when recording video responses on Recap) ...follow a theme and have like a whole coherent idea...where you have to speak more fully and really give a full complete answer.	

The Quality of the Speech is Enhanced

From excerpts (a) to (k), it is obvious that by replaying and re-recording the videos, the participants intentionally improved the quality of their L2 speech, specifically their fluency, accuracy, complexity, and content organization. At the same time, they were aware of and recognized this effort. Therefore, the participants' self-awareness and self-correction of their L2 were enhanced during this process. When planning and producing speech, the delayed nature of asynchronous communication provided the students with enough time to prepare and develop their speech in advance. By checking for accuracy, practicing for fluency, organizing and re-organizing ideas, using more complete sentences, and intentionally applying the newly learned vocabularies and grammar patterns, the general quality of the speech can be improved. These findings regarding presentational speaking practice are consistent with Wilches (2015)'s study.

Furthermore, as participant AW indicated in excerpts (l) and (m), the presentational nature of recording and posting a two-minute (or longer) video response required the speech to be delivered in a relatively more complete, coherent, and focused way, which is distinguished from relatively short answers produced in typical classroom teacher-student interactions. This finding is significant because it suggests that some learners are aware of the limitations of practicing speaking in class, and they are willing to practice presentational speaking using supplementary materials outside of the classroom.

A Sense of Achievement is Gained

During the interview, participants AW, EC, and EM indicated that at the beginning, they felt nervous and reluctant to record the video, partially because they were new to the experience of speaking Chinese in front of a camera and watching themselves at the same time, and also because they were not confident about their speech. However, after trying to record the video several times, they felt more concentrated, engaged, and confident. After they finished the activity, they realized that “this is actually really helpful” (SD). If we had adopted this kind of practice through a whole semester, then students’ progress could have been visually recorded (as AW said, “if we were doing this on a regular basis you could see your improvement”), which could help students improve their self-confidence and gain a sense of achievement. In Hsu et al. (2008), the investigator was surprised to learn that 8 out of 20 participants showed their audioblogs to their families or friends without any encouragement from the instructor or the researchers, which may be possibly attributed to the sense of achievement the participants gained from completing the presentational speaking tasks.

Interpretive Listening

The participants expressed that watching others’ videos was a fun and helpful experience. They indicated that they learned from each other, were happy to see that others shared the same opinions as them, and were interested about novel or different thoughts.

Table 4. Interview Excerpts of Participants’ Perceptions Toward the Interpretive Listening Activity

Participants	Excerpts from the interview (in English) / Comments from online discussion (in Chinese)	Themes
EC	(a) ...what they have used that we’ve learned that I could have used as well.	Learn from each other
SD	(b) ...they use this pattern that we learned, or they use this word that we learned, like I should use that...	
AW	(c) ...some of the things they talked about in theirs I didn’t think about it, oh yeah, that’s right, that’s actually really true, and so we can...share ideas so that when we go to the (speaking) exam we have lots more to talk about.	
AW	(d) Comment on EC’s video: 我们都同意, 不可以直呼老师的名字。我不知道为什么我们的课本说是这样。 [We both agree that it’s not permitted to call teachers’ names directly. I do not understand why our textbook says so.]	Share the same or different opinions
EC	(e) It’s happy to see that other classmates feel the same way about our textbook than I do. We both think it’s wrong.	

MS	(f) It's kind of fun to go and watch their videos and just see the different responses, see the different things they talked about. Because we are in the same class, but our responses were just a little bit different, so it was interesting.	
----	--	--

Learn from Each Other

Excerpts (a) and (b) in Table 4 suggest that the participants learned from listening to each other's videos by noticing some vocabulary and grammar patterns others had used and by realizing that they can utilize them in their speaking as well. This finding is consistent with that reported in Gleason and Suvorov (2011), who demonstrated that learners can significantly benefit not only from listening to themselves, but also from listening to others.

In excerpt (c), AW indicated that this kind of activity could be implemented before the monthly speaking exam, as it is helpful in assisting students to enrich their speech production by absorbing inspiring ideas from others' responses.

Share Opinions

One lesson of the textbook explains that in the U.S., students can always call their teachers by their first name directly; however, participants EM, AW, EC, SD (e.g., excerpts (d) and (e) in Table 4) argued that this is not true especially in pre-university education contexts. The participants reported during the interview that it was interesting and encouraging for them to hear or see that their classmates held the same opinions as themselves and that they both disagreed with the textbook on this particular point.

This finding is pedagogically inspiring as it provides second language instructors with another promising possibility to conduct interpretive listening or reading practice: Instead of asking L2 learners to listen to or read what the instructors compiled for them—which may be considered boring if the learners are not interested in the content of the assigned materials—the learners themselves can also be valuable creators of their learning materials. Furthermore, instead of implementing interpretive listening or reading activities individually, instructors can add them into an integral online discussion activity. When watching peers' videos and reading peers' comments, curiosity and responsibility to react to the ongoing discussion motivated learners to understand their classmates' words, through which they may be presented with novel or different ideas. Sharing similar opinions can also create a sense of camaraderie and solidarity in the classroom. During this energetic process, the interpretive listening or reading task becomes

relatively more meaningful and engaging because of the existence of authentic communicative intentions.

Interpersonal Online Discussion through Video Responses and Textual Comments

Table 5. Participants' Online Comments and Interview Excerpts About Interpersonal Online Discussion

Participants	Excerpts from the interview (in English) / Comments from online discussion (in Chinese)	Themes
AW	(a) ...watching their videos actually helped me understand...she is really fluent or...she knows the material really well	Become more familiar with each other
SD	(b) It may give an opportunity to interact with classmates on a different medium...especially with whom I do not talk to in the class	
AW	(c) Comments on SD: 你的朋友很可爱哦！ [Your friend is so cute!] (SD's beautiful parrot was standing on her shoulder when she recorded the video.)	
EC	(d) Comment on WA: 我很喜欢听你说话。你说得非常好，我嫉妒你 :) [I really enjoy listening to your speech, you spoke very well, I'm jealous of you!]	Peer-affective support
EM	(e) Comment on SD: 你说得很有道理！ [Your opinion is very reasonable!]	
AW	(f) Comment on EC: 我的口语不如你！ [My speaking is not as good as you!]	
SD	(g) It's good to hear feedback not only from the teacher but also from your classmates.	
AW	(h) It is encouraging and there is a lot of positive feedback.	
EC	(i) I liked (to see others' comments) because they know what I am saying...	

Become More Familiar with Each Other

As shown by excerpts (a), (b), and (c) in Table 5, through the opportunity offered by online discussion, participants get to know more about their classmates, either about their language learning situations or about their personal lives (e.g., the relationship with their pets).

In the college educational context, instructors often see that students choose to sit in the same place in the classroom for the whole semester and only talk to a limited number of classmates who sit nearby. By integrating a Recap-supported online discussion activity into a language class, students get more chances to communicate with all of their classmates outside of class, which is considered significant for the purpose of building a language learning community.

Peer-Affective Support

Comments (d), (e), and (f) in Table 5 show that participants complimented the general speaking performance of their classmates in their posted comments. During the interview, participants AW, SD, and EC also specified that it was encouraging to read others' comments and compliments of their performance in the videos, as exemplified by excerpts (g), (h), and (i). Obviously, this kind of affective support among peers can increase students' confidence and motivation toward L2 learning. This finding is also in line with previous studies.

Pedagogical Suggestions

Based on participants' general perceptions toward this activity as a whole and their constructive reflections, the following pedagogical suggestions were summarized for L2 instructors' reference.

Table 6. Interview Excerpts of Participants' Perceptions Toward the Whole Activity

Participant	Excerpt	Theme
EM	(a) I did, the feedback you gave me I did you know take look at what I have done incorrectly and so that was helpful.	Importance of giving individualized feedback
EC	(b) ...most time in class like the teacher would know what I'm saying but you won't necessarily correct me if I'm wrong.	
AW	(c) I do appreciate that you took the time go through each of ours and listen to everything.	
EM	(d) If I hadn't already known what they were talking about, it would have been pretty difficult, but since I already knew the topic, then that made it easier...	Importance of using class-related learned topic
AW	(e) It was good for us to recall the vocabulary...and practice with that vocabulary...if it was a new topic altogether it might have been more difficult and maybe not as helpful to use it in class because we are not working on that, so I think this was good to use the topic from class.	
EC	(f) I like to have your suggestions in the email like "我同意" [I agree] or "我不同意" [I do not agree] kind of thing. So like prompts, cause otherwise, I may not really know what to say.	Importance of providing language prompts

Giving Individualized Public Feedback

As some previous studies have documented (e.g., Hsu et al., 2008; Rashed, 2009; Xu & Peng, 2017), individualized feedback is considered crucial in CALL activities. Not surprisingly, this finding was also supported by the present study. All the participants confirmed that they had read the feedback carefully, and they emphasized the significance of receiving individualized feedback. Participants AW, EC, and EM even mentioned it several times during the interview.

Feedback was considered important by the participants (see excerpts (a) and (b) in Table 6) as it helped improve speaking accuracy. Furthermore, the feedback offered by the instructor through Recap effectively supplemented the limited chances of giving detailed feedback to every student during class time.

Nevertheless, despite being worthwhile and generally appreciated, for a relatively large class, giving detailed feedback to every student's video may become a time-consuming task for the instructor. To address this problem, there are some platforms that are similar to Recap, but more powerful in that they allow instructors to record video feedback, such as FlipGrid (<https://flipgrid.com/>). Under the assistance of this web-based platform, the feedback process is expedited (as it is communicated orally as opposed to typed). FlipGrid also allows the feedback to be delivered in a more effective way in that the instructors can correct pronunciation directly through the video, which is hard to realize using textual comments. Furthermore, for learners with relatively high language proficiency, the instructor may also consider providing the feedback in the target L2 for the sake of maximizing the chances for learners to practice listening, as AW constructively suggested during the interview: "then we can hear how it could be done, and not just read it, because reading, a lot of us are better at reading than we are at listening and speaking, so maybe hearing it as well. It would be helpful."

Encouragingly, during the interview, all the participants said that they found it helpful, not embarrassing, to see the feedback; EC and AW even indicated that they had learned more by reading the feedback left for others. Accordingly, it might be pedagogically practical to supply individualized feedback with public access. Additionally, participants AW, EC, EM, and SD also suggested that besides offering feedback, the instructor is also very welcome to participate in the discussion and share his/her personal opinions about the topic. Again, instructors may also consider recording a video response instead of typing comments depending upon the specific teaching context.

Using Expanded Class-Related Topic

Excerpts (c) and (d) in Table 6 represent the participants' shared perceptions that they preferred to discuss a class-related topic compared to a topic that is not related to the content learned on class. One benefit of presentational speaking practice is the opportunity to apply newly learned vocabulary. In Recap, this advantage also transfers to interpretive listening, as learners can

understand better what other classmates said because they are sharing the common knowledge of newly learned vocabulary and structures. Otherwise, the discussion would probably not be developed effectively. Additionally, choosing a class-related topic bridges the after-class practice with the content currently learned in class. This way, the learners would feel that their work is purposeful and worthwhile.

Nevertheless, there are limitations to using a class-related or textbook-based topic as it may lead to an unfavorable situation in which students simply repeat what is in the textbook instead of expressing their own opinions, as MS's representative comment indicated: "a lot of the conversations were just about what was in the textbook, that wasn't their opinion." To avoid this undesirable result, the selected topic is suggested to be both based on the textbook, as well as beyond the textbook. For example, during the interview, EM gave some very constructive suggestions:

If there is something (in the textbook) that you (the instructor) know probably isn't right then you could just ask the students 'Do you think this is really true?'...if they were to give their own opinion, then you would have something to say for backup like 'Oh, well in China it's like this...' and 'How do you guys do this?' Still using the same kind of grammar and all that, words we're familiar with, but just kind of generating a little more, follow up.

Based on her reflection, some possible sentence patterns for constructing an effective question that motivates students to express their original opinions and facilitates in-depth discussions are: "In the textbook, it said..., but actually in China (or any other applicable L2 speaking countries)..., so what is your opinion about this issue?"; or "In the textbook, it said..., do you agree with this view? Why or why not?" This kind of question encourages students to think about the related issues more critically. In this way, more meaningful and insightful conversations can be generated, facilitating an in-depth discussion.

After posting the question, to assist participants in expressing themselves efficiently in the target L2 during the online discussion, offering language prompts are considered necessary and helpful, just as EC in excerpt (e) (Table 6) mentioned.

Conduct the Activity as an Assignment

All the participants confirmed that this Recap-supported online discussion activity was very useful and beneficial (e.g., (a) and (c) in Table 7); therefore, it is considered worthwhile to make

the activity a graded assignment or one that counts towards the students' class participation score (see excerpts (b) and (c) in Table 7). This way, students will treat it more seriously by putting in more effort to complete the activity and, as such, learning is further supported and encouraged.

Table 7. Interview Excerpts of Participants' Suggestions for Making the Activity a Graded or Participation Assignment

Participants	Excerpts	Theme
AW	(a) It was good exposure, and I think that is one of the things we all struggle with, this listening and speaking. I try to go out and find podcasts and other things to do, but this it just makes me speak. And you know, if I go online and find somebody like a pen pal or something, it's hard to figure out OK what do we talk about and then how much they going to actually correct me, you know. And how much they're just going to let me talk in broken and try to figure out what I am saying. So, this is really a focused way to practice speaking and then get feedback, so it's good. (b) This could be an assignment...maybe even not like a graded assignment but just a participation assignment where you just get five points for doing it.	It's considered important to treat this activity seriously and assign it as an assignment
EC	(c) I think it's very useful and it is actually very beneficial. So, I think it would be most useful if you use it in the classroom to have some points associated with it...not just be like a minimal thing because otherwise people might not take it seriously...for this maybe if it wasn't like a real assignment I wouldn't have typed like more responses, you know. Otherwise, people might just not really put that much effort into it. But yeah, I think like if people want to take it seriously, it's actually really useful.	

Conclusion

In this study, the researcher investigated five L2 Chinese learners' perceptions of interacting with Recap (a web-based video prompt online discussion platform) by completing an online discussion activity, in which three communicative modes (presentational, interpretive, and interpersonal) were integrated. All the participants expressed a very positive attitude toward both the platform (Recap) and the activity.

In terms of practicing presentational speaking, through the process of recording, playing back, making self-corrections, and re-recording, the overall quality of the posted speech can be improved, as evidenced by the increased focus on fluency, accuracy, complexity, and content organization. The participants became aware of the areas where they struggle and intentionally worked to improve upon these aspects. Additionally, learners (with relatively high language proficiency) are aware of the necessity of practicing presentational speaking, which demands a

more complete, coherent, and focused oral production compared to relatively short and incomplete conversational turn-takings. Platforms such as Recap can visually record learners' progress longitudinally, which is a pedagogically meaningful way to build and enhance learners' sense of achievement.

In terms of practicing interpretive listening, by carefully listening to others' videos, learners can learn from each other's vocabulary and sentence structure, as well as share similar, different, or novel opinions. More importantly, the authentic communicative characteristic of Recap permits listening materials such as learners' self-recorded videos to make listening practice more engaging and meaningful compared to more traditional listening materials assigned by the instructor.

Recap also supports interpersonal online discussion through video responses and textual comments. It facilitates the formation of a class-based language learning community, as it creates more after-class communication opportunities for L2 learners to become familiar with each other and get to know more about each other. Students can also support each other by sharing positive feedback. The ongoing development of this community greatly benefits learners on their journey of learning a second language.

The participants also shared pedagogical suggestions based on the general perception of the activity as a whole. All participants appreciated the significance of getting individualized feedback from the instructor on their video responses. They also welcomed the instructor to participate in the discussion and would have liked to see the instructor's personal opinions surrounding the topic in question. The participants approved the appropriateness of using a class-related topic in an online discussion activity, while they indicated that they were willing to see the question constructed in a more debatable way to facilitate the generation of more exciting conversations. Based on their awareness of the advantages of this Recap-supported online discussion activity, participants suggested making this activity a graded or participation assignment, to facilitate all students to take it seriously and complete it carefully.

Limitations

One limitation of this study is the relatively small number of participants, which may underrepresent learners' perceptions in general on a larger scale. A second potential limitation is

the online discussion activity was only conducted once in this study, and because all the participants were new to Recap, it might be possible that their generally positive attitude toward Recap and the activity could be partly attributed to their novel feeling about this kind of platform. Accordingly, future studies should implement additional similar online discussion activities more and observe whether learners' perceptions evolved longitudinally, and if so, explore how and why they may have changed.

References

- Abrams, Z. I. (2003). The effect of synchronous and asynchronous CMC on oral performance in German. *Modern Language Journal*, 87(2), 157–167.
- ACTFL. (2012). *ACTFL proficiency guidelines 2012*. Alexandria, VA: Author.
- Beauvois, M. H. (1994). E-talk: Attitudes and motivation in computer-assisted classroom discussion. *Computers and the Humanities*, 28(3), 177–190.
- Blake, R. J. (2017). Technologies for teaching and learning L2 speaking. In C. A. Chapelle & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 107–117). Hoboken, NJ: John Wiley & Sons.
- Conole, G. (2008). Listening to the learner voice: The ever changing landscape of technology use for language students. *ReCALL*, 20(2), 124–140.
- Coverdale-Jones, T. (2000). The use of video-conferencing as a communication tool for language learning: Issues and considerations. *IALLT Journal of Language Learning Technologies*, 32(1), 14.
- Davin, K., Troyan, F. J., Donato, R., & Hellman, A. (2011). Research on the Integrated Performance Assessment in an early foreign language learning program. *Foreign Language Annals*, 44(4), 605–625.
- Elola, I., & Oskoz, A. (2010). Collaborative writing: Fostering foreign language and writing conventions development. *Language Learning & Technology*, 14(3), 51–71.
- Fazal, S., Majoka, M. I., Khan, M. I., & Masood, S. (2017). Technologies supported communicative grammar translation model: A motivational agent for English learning. *Pakistan Journal of Psychological Research*, 32(1), 175–193.
- Gilbert, J. (2017). A study of ESL students' perceptions of their digital reading. *The Reading Matrix: An International Online Journal*, 17(2), 179–195.
- Gleason, J., & Suvorov, R. (2011). Learner perceptions of asynchronous oral computer-mediated communication tasks using Wimba Voice for developing their L2 oral proficiency. In S. Huffman & V. Hegelheimer (Eds.), *The role of CALL in hybrid and online language courses* (pp. 1–12). Ames, Iowa: Iowa State University.
- Hsu, H.-Y., Wang, S.-K., & Comac, L. (2008). Using audioblogs to assist English-language learning: An investigation into student perception. *Computer Assisted Language Learning*, 21(2), 181–198.
- Lee, L. (2004). Learners' perspectives on networked collaborative interaction with native speakers of Spanish in the U.S. *Language Learning & Technology*, 8(1), 83–100.
- Lee, L. (2005). Using web-based instruction to promote active learning: Learners' perspectives. *CALICO Journal*, 23(1), 139–156.
- Levy, M. (2015). The role of qualitative approaches to research in CALL contexts: Closing in on the learner's experience. *CALICO Journal*, 32(3), 554–568.

- Levy, M., & Steel, C. (2015). Language learner perspectives on the functionality and use of electronic language dictionaries. *ReCALL*, 27(2), 177–196.
- Li, L. (2016). CALL tools for lexico-grammatical acquisition. In F. Farr & L. Murray (Eds.), *The Routledge handbook of language learning and technology* (pp. 461–477). New York, NY: Routledge.
- Li, M., & Storch, N. (2017). Second language writing in the age of CMC: Affordances, multimodality, and collaboration. *Journal of Second Language Writing*, 36, 1–5.
- Lin, L.-F. (2016). The impact of video-based materials on Chinese-speaking learners' English text comprehension. *English Language Teaching*, 9(10), 1–13.
- Lin, S. M., & Griffith, P. (2014). Impacts of online technology use in second language writing: A review of the literature. *Reading Improvement*, 51(3), 303–312.
- Litman, D., Strik, H., & Lim, G. S. (2018). Speech technologies and the assessment of second language speaking: Approaches, challenges, and opportunities. *Language Assessment Quarterly*, 15(3), 294–309.
- Ma, J. (2006). *Second language learners' perceptions of email communication in English* (Unpublished master's thesis). Queen's University, Kingston, Ontario.
- Nuessel, F., & Marshall, A. D. (2008). Practices and principles for engaging the three communicative modes in Spanish through songs and music. *Hispania*, 91(1), 139–146.
- Pardo-Ballester, C. (2012). CALL evaluation: Students' perception and use of LoMásTv. *CALICO Journal*, 29(3), 532–547.
- Rashed, D. H. (2009). *A case study of international ESL learners' perceptions of technology use in English language learning* (Unpublished master's thesis). University of Maryland, Baltimore, Maryland.
- Sotillo, S. M. (2000). Discourse functions and syntactic complexity in synchronous and asynchronous communication. *Language Learning & Technology*, 4(1), 82–119.
- Swanson, P. B., & Schlig, C. (2010). Improving second language speaking proficiency via interactional feedback. *International Journal of Adult Vocational Education and Technology (IJAVET)*, 1(4), 17–30.
- Tedick, D., & Cammarata, L. (2006). *Integrated performance assessment: Adapting the model for CBI*. Retrieved from <https://carla.umn.edu/cobalitt/modules/assessment/ipa/index.html>
- Wang, L., & Coleman, J. A. (2009). A survey of Internet-mediated intercultural foreign language education in China. *ReCALL*, 21(1), 113–129.
- Wang, S., & Smith, S. (2013). Reading and grammar learning through mobile phones. *Language Learning & Technology*, 17(3), 117–134.
- Warschauer, M. (1996). Comparing face-to-face and electronic discussion in the second language classroom. *CALICO Journal*, 13, 7–26.
- Yang, S. C., & Chen, Y.-J. (2007). Technology-enhanced language learning: A case study. *Computers in Human Behavior*, 23, 860–879.

“Her Russian Broke my Brain”: The Linguistic Influence of English on the Vocabulary of a Russian Heritage Speaker

Janet Stewart, University of Minnesota

The present study is an interlanguage analysis that explores the linguistic influence of English on the vocabulary of a Russian heritage speaker, with specific attention paid to the domain of concrete and abstract nouns. The interlanguage data for this analysis come from a popular Russian vlogger on YouTube who lives in Canada and predominately speaks her second chronological language (L2) English, but produces videos in Russian, her heritage language and chronological first language (L1). While L1 influence on L2 has been studied extensively in the field of second language acquisition, the reverse effect of L2 influence on L1 is less well understood. As such, the present study aims to shed light on the patterns of L2 influence on L1 on the lexicon by exploring the domain of concrete and abstract nouns, probing more specifically into patterns in object naming of common household items. Data for this study is taken from two YouTube videos and is analyzed according to Pavlenko’s (2003) framework for L2 to L1 transfer in the lexicon and semantics, which tracks four main phenomena: lexical borrowing, loan translation, semantic extension, and lexical retrieval issues. The findings of this study suggest that vocabulary pertaining to common household items is subject to more cross-linguistic influence than one may assume.

As language teachers and learners, it is sometimes assumed that certain lexical and semantic domains may be easier to acquire than others. It only seems logical that vocabulary pertaining to common household objects, for example, would be acquired to a fuller degree than less common vocabulary about abstract concepts. For heritage language learners especially, who often predominately gain linguistic input in their first language (L1) from their families in household contexts, it seems safe to assume that the vocabulary used at home would be solidly acquired and thus less likely to be influenced by their second language (L2). This paper aims to investigate this assumption, exploring the object naming patterns of a Russian heritage speaker whose L1 Russian provides evidence of significant L2 English influence. In analyzing two YouTube videos in which the speaker draws on concrete and abstract ideas and language to describe how and where she lives, this study aims to shed light on the patterns of L2 influence on L1 on the lexicon, with specific emphasis on patterns in object naming of common household items in comparison to abstract nouns. The paper poses the following questions: To what extent are there

linguistic influences of English on the language of a Russian heritage speaker in object naming? Are there discernable patterns of L2 transfer in L1 naming of common household objects in comparison to abstract nouns?

Literature Review

The inspiration for this topic came from Pavlenko's (2011) *Kitchen Russian: Cross-Linguistic Differences and First-Language Object Naming by Russian-English Bilinguals*, in which it was discovered that the L1 Russian naming patterns of common household entities were influenced by the speakers' L2 English. The influence was strongest for early and childhood bilinguals, or heritage speakers, which is the term this paper will use to describe a speaker of a *heritage language*, an individual's first chronologically acquired language, which may have not been "completely acquired because of the switch to another dominant language. An individual may use the heritage language under certain conditions and understand it, but his/her primary language is a different one" (Valdes, 2000, as cited in Polinsky, 2008, p. 149). That being said, while heritage speakers, such as the speaker whose interlanguage is analyzed in the present work, may not have acquired their L1 to the same extent as a monolingual of that language, they have developed an interlanguage system in their own right, one that Cook (1992), in accordance with his multi-competence perspective, argues may be the product of the merging of their two language systems. In other words, an individual's L2 and L1 are seen as part of the same linguistic system and can thus interact within it. In adopting this perspective, we are able to analyze the influence of one language on another within an individual's linguistic system. More specifically and for the purposes of this paper, it provides us with the theoretical framework to look at linguistic influence of L2 on L1 lexicon.

According to Pavlenko (2011), L2 influence on L1 may affect different lexical and semantic domains in different ways. One such domain, which offers great opportunity to observe L2 influence on L1 on lexicon, is the naming of common household objects. As long as the heritage language continues to be used at home, the domain of household objects should remain consistent for immigrant children and heritage speakers. Additionally, the objects in this domain are familiar, commonly used items, and thus, are likely to be an established part of a heritage learner's acquired lexicon. Nevertheless, while it seems logical that common household items, a

familiar and frequently accessed lexical domain, would not be subject to L2 influence, the opposite was determined to be true. In naming various types of drinking containers, for example, heritage speakers tended to simplify the object's name and showed a general shift toward English object naming trends. To illustrate, when provided pictures of various clear, glass drinking containers (*kruzhka* [mug], *bokal* [goblet], *riumka* [wine glass], etc.), the majority of heritage speakers named the object *stakan*, a tall, cylindrical glass cup, which has a translation equivalent in English of a generic drinking glass. English naming conventions tend to focus on material, whereas a Russian naming trend is to focus more on size and shape. Thus, childhood bilinguals in this study drew on their L2 English lexical systems to help them name L1 Russian objects, despite what could be considered a familiar domain. Pavlenko explained why this trend may occur, suggesting that “naming patterns for these common objects in one language can be swayed by exposure to patterns in another just as use of more complex or abstract language is” (p. 39). From this, it can be understood that common household objects are, indeed, subject to cross-linguistic influence; however, we do not gain a perspective on the degree to which this influence happens, or how the influence on household objects compares to the influence on a domain such as abstract nouns.

Apart from Pavlenko (2011), very few studies exist that specifically explore the influence of L2 on L1 lexicon at the level of a particular noun category. There is, however, evidence to support that the lexical processing of abstract and concrete nouns differs, especially in regard to second language acquisition. In fact, it is widely considered that L2 learning verbal processes evolve from the concrete to the abstract (Mägiste, 1984), meaning that a difference must exist. In addition, it has been shown that bilinguals have greater recall and recognition for concrete words than for abstract (Mägiste, 1979; Winograd, Cohen, & Barresi, 1976). In addition to being recalled faster, more recent findings suggest that bilingual speakers also translate concrete nouns, which are more likely to have only one unique translation equivalent, faster than abstract nouns (de Groot et al., 1994; van Hell & de Groot, 1998). Lastly, in a study that compared word associations made by monolingual English and Spanish speakers with English-Spanish bilinguals living in Spain, Grabois (2000) provided further evidence for a distinction in abstract and concrete noun processing. Through multiple word association tasks, it was discovered that “lexical networks of a series of abstract concepts, including power, love, and happiness, were

different and relatively consistent for the two monolingual groups” (Pavlenko, 2000, p.188); however, the English-Spanish bilinguals’ word associations were closer linked to those of L1 Spanish speakers, despite being their L2. What this study and past research suggests is that abstract lexical items are generally more difficult and slower for bilingual speakers to access. In response to the challenge of naming abstract concepts, bilingual speakers may draw on their L2 to provide lexical support. In the process of drawing on one’s L2, cross-linguistic influence is inevitable.

A review of past research suggests a gap in knowledge about L2 to L1 influence on lexical domains of concrete and abstract nouns. The results of Pavlenko’s (2011) *Kitchen Russian* raise a particular question, which is if L2 to L1 influence in the mental lexicon can occur in a domain involving concrete nouns in reference to something familiar, such as common household objects, how does this influence compare for abstract nouns, which, as research shows, are a more difficult lexical item to access? In other words, what this study aims to answer are the following questions:

1. To what extent are there linguistic influences of English on the language of a Russian heritage speaker in object naming?
2. Are there discernable patterns of L2 transfer in L1 naming of common household objects in comparison to abstract nouns?

Methodology

The Data

The interlanguage data for this analysis come from a popular Russian beauty vlogger on YouTube named Sonya Esman. Born in Russia, she is a Russian heritage speaker, who moved to Canada at the age of five. At the time that the analyzed videos were posted, Sonya was sixteen and still living in Toronto. In this paper I will consider her L1 to be her chronological first language acquired, Russian. Her dominant language, North American English, will be referred to as her L2. In answering questions about her language use, she claims that her only interactions in Russian are with her mother, a native Russian speaker from Voronezh, Russia, and on her Russian YouTube channel, which at the time of the present study has 1.8 million subscribers. According to the information she reveals in her vlogs, she has not had any formal instruction in Russian, and she does not express interest in further developing her language.

The Videos

It is common for vloggers on YouTube to assign themselves a specific topic to discuss or task to complete in their videos. Video ideas frequently come as requests from subscribers and commenters. In producing content based on these requests, many videos become a type of task. “Life in Canada” (henceforth referred to as Video 1), for example, takes the form of a comparative task. In answering questions from her subscribers about life in Canada, Sonya ends up comparing aspects of life in Canada, Russia, and sometimes the U.S. This entire video is shot in true vlogger-style, showing only herself on screen, so she has no visuals other than her own body and facial expressions to convey meaning while referencing various items and ideas. “My Room in the French Style” (Video 2) is a tour of her redecorated bedroom and takes the form of a descriptive task.

I picked two videos that I believed would produce opposite nouns categories. For example, I chose Video 2 because it requires the vlogger to talk about physical objects that are shown on screen. Because of the visual nature of the room tour, my assumption was that Video 2 would produce mostly concrete nouns, whereas the country comparison in Video 1, which lacked any identifying objects, may necessitate the use of more abstract nouns. In looking at the naming of concrete nouns and common household items, having a visual for verification adds a beneficial layer for analysis.

Lastly, I chose videos posted between 2011 and 2012 because they are some of the least edited, and thus most authentic examples of Sonya’s spontaneous Russian use. There are no special video effects or additional elements. In later videos from more recent years, for example, Sonya occasionally overlays mispronounced or misused words as text on the video in post-edits. For the purposes of this study, I wanted to analyze what was said in the moment of filming. A summary of the videos I chose to analyze can be seen in Table 1, below.

Table 1. Summary of Videos

	Video 1	Video 2
Title	Zhizn' v Kanade <i>Life in Canada</i>	Moya komnata v frantsuzskom stile <i>My Room in the French Style</i>
Date posted	November 11, 2011	September 12, 2012
Duration	9:50 minutes	9:04 minutes
Topic	Country comparison	Room tour
Visual	Sonya speaks directly to the camera; her face and kitchen wall are the only visuals.	Sonya speaks directly to the camera in the beginning and end; she shows objects in her bedroom for the majority of the video.

Analysis

To describe Sonya's vocabulary, I looked at lexical complexity, abstract and concrete noun use, and L2 effects on L1 lexis. First, I transcribed the videos as they were in Russian. After checking the transcriptions against each video several times, I translated the transcripts into English and organized them into tables.

To measure lexical complexity, I calculated the type-token ratio (TTR) for both videos. This was done to gauge the speaker's overall lexical complexity as well as to ascertain whether or not any large differences in complexity between the two videos could be found, as that could potentially impact the later parts of my analysis. I selected one segment from the first video at random and looked at lines 38–47. For Video 2, I looked at a section of lines at about the same point in the transcription: lines 36–46. I then had to determine my criteria for tokens. I removed universal fillers, such as *um*, and *uh*, as well as the Russian-specific filler words *nu* [well], *tam* [there], *da* [yes]. I considered *nu da* [so yeah] as one token, because it is not used as a filler as much as a way to transition from one topic to another. In fact, there were a couple of two-word Russian transitions and adverbs, such as *Tak chto* [so that] and *potomu chto* [because], that I counted as one token based on semantic values. Lastly, I considered imperfective and perfective verb pairs as one token, despite sometimes appearing as separate forms (*skazat'* and *govorit'*, for example, are the perfective and imperfective forms, respectively, of the verb "to say"). Once the criteria for tokens were determined, they were counted. Types, single instances of a word, were then counted and compared over the total tokens.

Next, to analyze concrete and abstract noun use, I pulled out every noun, common and proper, and categorized them as concrete or abstract, maintaining their organization by task. Pronouns and repeated words were not included. In addition, due to frequent misuse, which at times caused a shift in meaning, diminutive forms were counted as separate words from the head nouns (e.g., *komntata* [room] and its diminutive *komnathka* [little room]). I defined abstract nouns as ideas, conditions, and qualities, as well as things that cannot be seen or touched and have no physical reality, such as units of time, natural elements, and the Internet. Concrete nouns were defined as people and things that exist physically and can be seen, touched, smelled, heard, or tasted, including physical spaces such as countries and stores. Once the nouns were categorized, I determined whether or not each concrete noun could be considered a common household item. Common household items were determined on the basis that they were: (1) specific, concrete objects, (2) typically found within a house, and (3) a common or familiar item. For example, nouns such as bed or table were counted, whereas nouns such as *veshi* [things] were not considered on the basis of being too general. I then tallied the number of nouns by type, task, and total, the results of which will be discussed in the next section.

Finally, I looked at L2 influence on L1 in the lexicon and semantics, adopting the framework used in the analysis of Pavlenko (2003), in which she outlined four main phenomena:

1. *Lexical borrowing*, or the adoption of L2 items into the L1 lexicon morphologically or phonologically. *Boifrend* [boyfriend] or *dauntaun* [downtown] are well-documented examples of this phenomena among Russian immigrants (Andrews, 1999).
2. *Loan translation*, or the literal translation of L2 words, compound words, expressions, idioms, or lexical collocations. *Neboskreb* [skyscraper], for example, is a literally translated compound of *nebo* [sky] and *skreb* [scrape] (Pavlenko, 2000).
3. *Semantic extension*, or loan shift, is an extension of the use of L1 lexical items to include the semantic values of an assumed L2 translation equivalent (Otheguy et al., 1989). Pavlenko (2003) subdivided it further into the following categories:
 - a. Assumed translation equivalent;
 - b. Semantic narrowing, or loss of semantic distinction;
 - c. Inappropriate register of references; and
 - d. Attribution of a particular meaning to a word based on superficial word-structure similarities.

4. *Difficulties in lexical retrieval.* In my analysis, I considered instances only when the speaker explicitly stated difficulty or uncertainty in expressing herself.

Using this as my framework, I performed a lexical error analysis, in which I looked at each noun in context and assessed the accuracy of the term. If I determined it to be an error, I brainstormed possible reformulations or hypotheses for the error. If the error seemed to be influenced by English, I used the framework stated above to determine the effect type. Apart from personal knowledge, I drew on several other resources to guide this process: an online Russian dictionary and encyclopedia database, which generates definitions from many sources for comparison; two online translators run by major search companies; and the Russian National Corpus, which incorporates over 300 million words from modern Russian language.

Findings

Overall Lexical Complexity

The TTR analysis showed a consistent rate of lexical complexity in both videos. For instance, the segment from Video 1, which contained 110 tokens, consisted of 67 types, meaning there was a TTR of .61. In comparison, Video 2, which contained 112 tokens, 62 of which were types, produced a TTR of .55. The results of this analysis show that the lexical complexity between the two videos, despite being about different topics and in different formats, was fairly consistent.

Table 2. Type to Token Ratio for Both Videos

	Types	Tokens	Type-token ratio
Video 1—Comparison	67	110	0.61
Video 2 —Room Tour	62	112	0.55

Concrete and Abstract Nouns

Next, I analyzed the numbers of concrete and abstract nouns, also singling out common household items. First, the final count of total unique nouns between both videos was surprisingly even, with 97 total unique nouns in Video 1 and 100 in Video 2. Combined, there were a total of 197 nouns; however, 21 of those nouns occurred in both videos. Thus, looking at the complete dataset, there were 176 total unique nouns to analyze.

Overall, both videos contained more concrete than abstract nouns; however, the videos did not produce them at an even rate. In Video 1 65% of the nouns were concrete, whereas in

Video 2, the room tour, 78% of the nouns were concrete. Video 1, the country comparison, contrarily, contained more abstract nouns than Video 2, with a difference between 35% and 22%, respectively. These data, as shown in Table 3, support my hypothesis that the comparison video would necessitate more abstract nouns than the room tour. In looking at the distribution of unique concrete and abstract nouns of both videos combined, 72% of the unique nouns are concrete, while 28% are abstract.

Table 3. Unique Concrete and Abstract Nouns per Video

	Video 1	Video 2	Both
Concrete nouns	63 (65%)	78 (78%)	127 (72%)
Abstract nouns	34 (35%)	22 (22%)	49 (28%)
Total unique nouns	97	100	176

Of the 127 unique concrete nouns, 48 counted as common household items, meaning that nearly 38% of all concrete nouns fell into this category, and that common household items accounted for 27.2% of total unique nouns. In the next section I will discuss the results of how each of these noun categories were influenced by L2 English.

Lexical and Semantic Errors in Object Naming

Altogether, I found thirty instances of L2-influenced lexical and semantic errors: fourteen instances of lexical borrowing, nine of semantic extension, four loan translations, and three lexical retrieval issues (see Table 4).

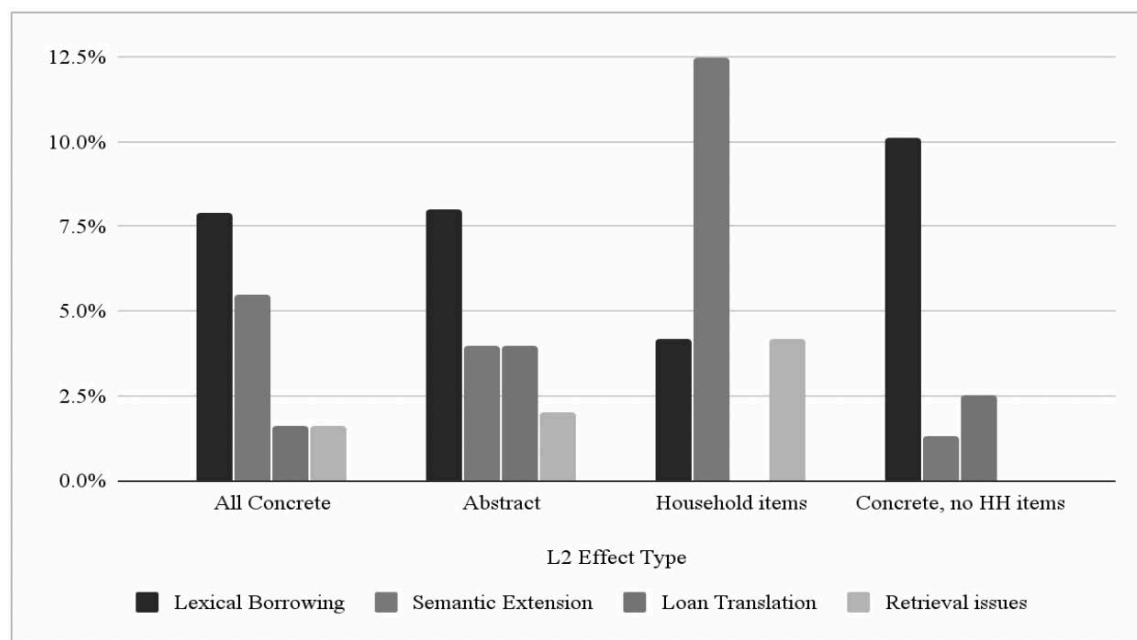
Seventy-percent of the L2-influenced errors occurred with concrete nouns, while 30% were with abstract nouns. Interestingly, this is the same distribution of abstract and concrete nouns in total (of all nouns analyzed, 72% were concrete, 28% abstract). Looking at the rate of influence, abstract nouns were influenced just barely more at 18.4%, while concrete nouns were influenced at a rate of 16.5%. This is likely not statistically significant, meaning that abstract and concrete nouns seem to be influenced at nearly equal rates. Also, as demonstrated by the order of effect type from most common to least, the total concrete and abstract nouns categories show a corresponding pattern, with lexical borrowing as the most common.

Table 4. Amount of Type of L2 Effect on Concrete and Abstract Nouns

L2 effect type	Total instances	Concrete nouns			Abstract nouns n = 49
		Household items n = 48	Concrete, no HH n = 79	All n = 127	
Lexical borrowing	14	2 (4.2%)	8 (10.1%)	10 (7.9%)	4 (8%)
Semantic extension	9	6 (12.5%)	1 (1.3%)	7 (5.5%)	2 (4%)
Loan translation	4	0 (0%)	2 (2.5%)	2 (1.6%)	2 (4%)
Retrieval issues	3	2 (4.2%)	0 (0%)	2 (1.6%)	1 (2%)
Total	30	10 (20.8%)	11 (13.9%)	21 (16.5%)	9 (18.4%)

As Figure 1 (below) depicts, the frequency pattern for L2 effect type that is shared by concrete and abstract nouns is not shared by noun categories affected by common household items (common household items and concrete nouns without household items). Whereas lexical borrowing affected the most nouns overall, common household items were affected most by semantic extension. Lexical borrowing and retrieval issues were the next common effects on household items, affecting them equally at a rate of 4.2%. Interestingly, when looking at concrete nouns, retrieval issues were exclusive to common household items. Also, there were no instances of loan translation among the L2-influenced common household items.

Figure 1: L2 Effect Type by Noun Category and Rate



In total, 20.8% of all common household items showed L2 influence. In comparison to all examined noun categories, this was the highest rate of influence (followed by abstract nouns at 18.4% and all concrete nouns at 16.5%). These results will be discussed in the next section.

Next, I illustrate some examples from the data by each type of L2 effect on L1 lexis. As previously discussed, nouns were analyzed according to Pavlenko's (2003) framework for L2 influence on L1 in the lexicon and semantics, which considers the following four phenomena: lexical borrowing, semantic extension, loan translation, and lexical retrieval issues. The results of my analysis of these phenomena are exemplified below.

Lexical Borrowing

Almost half of all instances of L2 influence on nouns were on account of lexical borrowing. Lexical borrowing affected the category of concrete nouns without household items most significantly, at a rate of 10.1%, and affected common household items the least, at 4.2%.

Some examples of lexical borrowing from the data include: *molly* [malls] (1-50),¹ *dauntaun* [downtown] (1-51), and *lanch* [lunch] (1-54). While all of these examples are used in modern Russian language, a comparison with the more normative counterparts in the Russian National Corpus shows that they are used far less often. *Moll* [mall], for example, occurs in 23 contexts, whereas *torgovyy tsentr* [shopping center] appears in 173 contexts. *Lanch* [lunch] appears in 156 contexts, whereas *obed* [lunch] occurs in 11,403 contexts. Lastly, *dauntaun* [downtown], which was cited by Andrews (1999) as a well-documented form of lexical borrowing, only occurs in eight contexts in the corpus, whereas *v tsentre goroda* [in the center of town] appears in 987 contexts. Thus, although these forms are seen in modern Russian, they are less likely to be used than other terms.

Another interesting example of lexical borrowing in Sonya's data was in describing two-pound dumbbells that she had sitting under her desk. Instead of the Russian word *funt* [pound], Sonya adopted the English word pound, changing the morphology several times in an attempt to match with the Russian case system.

(2-45) *Tam u menya gantel'ki na dva **pound-iy**. Ya sprosila mamu kupit' mne gantel'ki*

¹ Notation for examples is as follows: (video number - line number)

[There I have little dumbbells for two **pounds**. I asked my mom to buy me little dumbbells]²

- (46) *ona prinesla mne dva pound-a. Chto ya mogu delat' s dva pound-ami?*
[she brought me two **pounds**. What I can do with two **pounds**?]

In Russian, the numbers two through four are followed by the genitive singular case, which is marked with the *-a* or *-ya* suffix for masculine and neuter nouns, and *-i* or *-iy* ending for feminine. From the example above, it appears that Sonya had not established the gender of her borrowed term. In the first instance, she treated it as feminine, adding *-iy* to the end; however, the second instance is also followed by *dva* [two] and she used the masculine *-a* ending. In the final instance, she used the instrumental case and correctly added the *-ami* ending; however, this ending is the same for all genders. In this example, it seems that Sonya assumed that a cognate for pound exists in Russian.

Semantic Extension

Semantic extension was the next common overall L2 effect type, occurring in nine of 30 instances. Semantic extension was the most common L2 effect type for common household items. There were six instances when semantic extension involved a common household item, such as *kamera* [chamber], which was used to express *fotokamera/fotoapparat* [camera], and *doska* [board], which was extended to express (*pis'mennyy*) *stol* [(writing) table or desk].

I further analyzed instances of semantic extension in accordance with the subdivisions in Pavlenko's (2003) framework. Of the four subdivisions, many of Sonya's instances of semantic extension involved making references in an inappropriate register. Most often, this meant an imprecise use of the diminutive. In Russian, the diminutive form has many functions and meanings. Despite the many ways to use it, Sonya tended to diminutivize nouns only to imply that something was either small or cute. Although this is a very common function of the Russian diminutive, this meaning is not guaranteed by the addition of a diminutive suffix. In many cases, adding a diminutive suffix, such as *-ik*, *-chik*, *-yok*, *-ichka*, *-ka*, etc., extends and even changes the meaning. The best example of that in Sonya's data is the word *znachok* [pin/badge], which she

² For the sake of analysis, translations here are presented in English as close to word-for-word order as was possible.

uses to describe a *znak* [sign]. In this case, adding the *-chok* diminutive ending changed the meaning to the point that had there not been visual representation of the object shown in the video, it would have been assumed to reference a completely different object. While other examples were not as extreme, they did change the reference. When showing her floor-length curtains, for instance, Sonya referred to them as *shtorki* which was formed by adding the diminutive suffix *-ki* to the appropriate word *shtory* [curtains/drapes]. *Shtorki* often translates to blinds, but also commonly refers to the short panels used to block light in car windows.

Loan Translation

Loan translation was the third overall most common L2 effect. There were only four instances, which were split evenly between abstract nouns and concrete nouns. There were no instances of loan translation with common household items. Instances of loan translations are literal translations of English expressions into Russian, as evidenced by the following example:

(1-43) *nado sdelat' sorok "volonterskikh chasov"* {air quote}
[you have to do forty "volunteer hours"]

In Video 1, Sonya discussed graduation requirements for her school in Toronto. One of those requirements is to complete forty volunteer hours. First, the concept of volunteering up until recently in Russia had been considered a foreign idea, as evidenced by the borrowed vocabulary used to discuss it (*volonter* [volunteer], for example, is a borrowed word from English). The reason volunteer hours was categorized as loan translation, and not as lexical borrowing, is due to the way Sonya combined the components. In the English phrase, volunteer is serving as an adjective in order to modify the noun hours. She seems to apply this logic in order to directly translate the phrase into Russian. *Volunter* [volunteer] takes on an adjectival ending *-skii* (in this case, the genitive plural ending *-skikh*) to become an adjective used to modify *chasov* [hours] (*chas* [hour] + *-ov* genitive plural ending). Thus, instead of borrowing an English word, she combined elements from both languages and produced a literal translation.

Retrieval Issues

There were only three clear instances of lexical retrieval issues. They were split between abstract nouns, which had one instance, and household items, which had two instances, one of which is presented below:

- (2-60) *ya sovsem zabyvayu kak eto nazyvayetsya nu, vnutri vot eti veshchi,*
[I completely forget what this is called hm, inside of these things,]
- (61) {pulls on duvet cover} *kak **podlyanik** mozhet byt' nazyvayetsya.*
{pulls on duvet cover} [like **podlyanik** maybe it's called.]
- (62) *Net eto-- blin! Nu, koroche vy ponimayete <laughter> o chem ya govoryu.*
[No it's-- shoot! Well, in short, you understand <laughter> what I'm saying.]

In this excerpt, Sonya is trying to find the target word for duvet cover, which is *pododlyanik* while the item is shown on screen. At first, she expressed difficulty accessing the term, saying she could not remember what it is called. Then, she produced what could be interpreted as a performance error: Instead of the target word *pod-od-lyanik* she produced *pod-lyanik* skipping a middle syllable. Despite her near-retrieval, she doubted the term and conceded, stating to the camera, “well, you understand what I’m saying.” As such, this example can easily be interpreted as lexical retrieval difficulty.

Discussion

The first research question posed in the study sought to determine the extent to which Sonya’s L2 English influenced her L1 Russian in object naming. Looking at the rate of L2 effects on all noun categories, the highest rate of influence occurred in the naming of common household objects, at 20.8%. In other words, object naming was subject to more linguistic influence from English than any other category. This is consistent with Pavlenko’s (2011) findings, which suggested that L2 to L1 lexical influence can occur even in domains involving something familiar, like common household objects.

The second part of the question aimed to discern any possible patterns of influence in the naming of common household objects in comparison to abstract nouns. First, it was found that when comparing concrete and abstract nouns, patterns for most to least common L2 effect type were the same. The pattern for common household objects, on the other hand, followed a different trend. For example, there were no instances of loan translation among the affected household items. In addition, amongst concrete nouns, retrieval issues were exclusive to common household items. This is surprising, considering that previous research suggests that

the retrieval of abstract nouns in lexical processing for bilinguals is slower and more difficult, and often relies on the L2 more to fill in (Mägiste, 1979; Winograd, Cohen, & Barresi, 1976).

The most common L2 effect for concrete and abstract nouns was lexical borrowing. Interestingly, lexical borrowing occurred at about half the rate for common household objects. Semantic extension, on the other hand, affected common household objects by a huge margin, where six of the nine instances of the phenomenon involved common household items, and only one and two involved concrete nouns with no household items and abstract nouns, respectively. These results beg the question: Why did semantic extension occur at such a high rate of influence on the naming of common household objects? One possible answer could be related to the use of the diminutive, which, as evidenced in the data, Sonya used as a generic way to imply the cuteness or small size of an object. It makes sense then, that when describing familiar household objects, Sonya would try to emphasize one of these characteristics. Further, according to the results of a survey conducted by Hasko (2010), in which 100 native Russian speakers aged 12–19 were asked whether they would use a list of nouns in the diminutive or not, it was found that “body parts, clothing, and common household objects used by children were overwhelmingly preferred in their diminutive forms” (p. 34). Thus, it is possible that Sonya’s extension of the diminutive in relation to the naming of common household objects is in part due to the nature of her L1 Russian acquisition. In other words, as someone whose natural L1 Russian acquisition was cut off at a young age by moving to an English-speaking community, Sonya’s pattern of diminutive extension may be part of a broader issue of developmental sequences. In returning to my original question, however, the results of this study conclude that object naming for household items in comparison to abstract nouns is not only subject to cross-linguistic influence, but also follows separate trends.

This was a preliminary and small-scale study that faced many limitations. Primarily, the determination of common household items was based on my own judgement and experience of what is and is not common in Russian and North American homes. As such, the data presented could be skewed to represent my own worldview. In the future, the study would be better served to have a more systematic methodology to make this determination. In addition, as a non-native Russian speaker, it is possible that in my analyses, examples of influence were missed or overstated. If I were to refine this study, I would incorporate the feedback of other Russian users.

Finally, it would be interesting to extend the analysis past common household items and compare patterns of L2 influence for more noun categories, possibly following Andrews' (1999) categorization of nouns in lexical borrowing and loan translations in the speech of Russian speakers. In addition to household items, he looked at noun categories of employment, the automobile, cuisine, daily life, and academics. Perhaps exploring a broader range of noun categories could lend itself to a deeper understanding of L2 influence on L1 lexicon.

Conclusion

By exploring the domain of concrete and abstract nouns, probing more specifically into patterns of L2 influence in object naming of common household items, this study sheds light on the patterns of L2 English influence on L1 on the lexicon of a Russian heritage speaker. Using the framework for analyzing L2 influence on L1 in the lexicon and semantics as set by Pavlenko (2003), this study illustrated differences in L2 effects on different noun categories, specifically illuminating the fact that vocabulary pertaining to common household items is subject to more cross-linguistic influence than one may assume.

As such, an important implication for teaching of this study might be to focus more on a heritage speaker's vocabulary development, acknowledging that certain terms, which may seem common and simple, may actually be subject to just as much, if not more, cross-linguistic influence as a different noun category. Because of this, no item should be swept over or assumed to be fully acquired. In general, further exploration in the field of heritage language vocabulary and lexicon would be beneficial to better understand the needs of heritage language learners in L2 classrooms.

References

- Andrews, D. (1999). *Sociocultural perspectives on language change in diaspora: Soviet immigrants*. Amsterdam, Netherlands: John Benjamins.
- Cook, V. (1992). Evidence of multicompetence. *Language Learning*, 42(4), 557–591. doi: 10.1111/j.1467-1770.1992.tb01044.x
- Esman, S. [Соня Есьман]. (2011, November 11). *Moya zhizn' v Kanade* [Video File]. Retrieved from <https://www.youtube.com/watch?v=igQsvD3wXZo>
- Esman, S. [Соня Есьман]. (2012, September 12). *Moya komnata v frantsuzskom stile* [Video File]. Retrieved from <https://www.youtube.com/watch?v=XL1Gu7IBJ0k&t=94s>
- Hasko, V. (2010). Diminutives in spontaneous narration by American learners of Russian. *The Slavic and East European Journal*, 54(1), 32–53.
- Mägiste, E. (1984). Stroop tasks and dichotic translation: The development of interference patterns in bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(2), 304–315. <http://dx.doi.org/10.1037/0278-7393.10.2.304>
- Otheguy, R., Garcia, O., & Fernandez, M. (1989). Transferring, switching, and modeling in West New York Spanish: An intergenerational study. *International Journal of the Sociology of Language*, 79, 41–52. <https://doi.org/10.1515/ijsl.1989.79.41>
- Pavlenko, A. (2003). “I feel clumsy speaking Russian”: L2 influence on L1 narratives of Russian L2 users of English. In V. Cook (Ed.), *Effects of the second language on the first* (pp. 32–61). Clevedon, England: Multilingual Matters
- Pavlenko, A., & Malt, B. (2011). Kitchen Russian: Cross-linguistic differences and first-language object naming by Russian-English bilinguals. *Bilingualism: Language and Cognition*, 14(1), 19–45. doi:10.1017/S136672891000026X
- Polinsky, M. (2008). Heritage language narratives. In D. Brinton, O. Kagan, & S. Bauckus (Eds.), *Heritage language education: A new emerging field* (pp. 149–164). New York, NY: Routledge.
- Winograd, E., Cohen, C., & Barresi, J. (1976). Memory for concrete and abstract words in bilingual speakers. *Memory and Cognition*, 4(3), 323–9. DOI: 10.3758/BF03213184
- Wong-Fillmore, L. (1991). When learning a second language means losing a first. *Early Childhood Research Quarterly*, 6, 323–346.

Appendix I Transcription Conventions

CAPS	emphasis
?	rising intonation
.	falling intonation indicating completion of thought
,	natural pause
--	abrupt pause or change of thought
!	exclamation
#	inaudible word or stretch of words
...	silence
//	IPA transcription
(words)	interpretation of questionable words
<laughter>	description of noises or preceding sound (laughter, accented, etc.) <i>Example: "I went to IKEA <Russian accent>" shows accent on the word "IKEA"</i>
{points}	nonverbal behavior, such as movements and looks
((notes))	notes by the transcriber
[words]	English translation
Bold	L1-Influenced noun

A Mixed Methods Evaluation of an Urban K–12 World Languages Program

Reuben Vyn, University of Iowa

In the current era of educational accountability, foreign language programs are under increasing pressure to provide evidence of student learning outcomes, often in the form of large-scale performance assessments (Malone & Sandroock, 2016). However, relatively little is known about the nature of student outcomes on such assessments or how these data are leveraged for the purpose of improving language teaching and learning, particularly at the K–12 level (Donato, Tucker, Wudthayagorn, & Igarashi, 2010). In response to calls for further research in this area (Malone, 2017; Sandroock, 2006), the current mixed methods evaluation study of an urban K–12 world languages program investigated secondary students' language performance outcomes, as well as the way that educators responded to and integrated such data into their practice. To that end, upper-level foreign language students' outcomes data from the American Council on the Teaching of Foreign Languages' (ACTFL) Assessment of Performance toward Proficiency in Languages (AAPPL) were complemented by survey and interview data, which addressed teachers' instructional and assessment-related practices. The synthesis of findings related to these complementary strands resulted in recommendations for program stakeholders, which were aimed at facilitating improvements to their instruction, curricula, and professional development opportunities.

Introduction

Although the field of foreign language education has not been directly implicated in policies such as No Child Left Behind (2002) or the Every Student Succeeds Act (2015), its stakeholders have nonetheless experienced the ripple effects of the burgeoning era of standardized testing and accountability (Cox, Malone, & Winke, 2018; Ricardo-Osorio, 2008). However, few studies to date have explored the ways that language performance assessment outcomes data have been leveraged for the purposes of facilitating innovation in language teaching and learning in the K–12 setting (Donato, Tucker, Wudthayagorn, & Igarashi, 2010). If our field is to take steps toward legitimizing its relevance among competing disciplines, it is important for foreign language educators and administrators alike to actively engage in the process of program evaluation, and to apply the knowledge produced therein for the purpose of improving students' proficiency outcomes (Kiely & Rea-Dickins, 2009). In response to calls for further research on student outcomes and the role of language performance assessments (Malone, 2017; Sandroock,

2006), the current study evaluated the performance of an urban K–12 world languages program by investigating (a) the extent to which students in an urban K–12 school district are meeting the program’s goal of achieving intermediate levels of proficiency prior to program completion, and (b) the ways that language educators responded to student outcomes data for the purpose of making improvements in their practices.

Literature Review

Program Evaluation

The term evaluation is frequently misunderstood or misrepresented (Norris, 2016). Perhaps this is because its design, purpose, data sources, and products are largely dependent upon the individual needs of stakeholders that request this unique approach to scientific inquiry.

Nevertheless, Norris’ (2006) definition will be used as a point of departure: “Evaluation is the gathering of information about any of the variety of elements that constitute educational programs, for a variety of purposes that include primarily understanding, demonstrating, improving, and judging program value” (p. 579). Thus, program evaluation is particularly apt at facilitating data-based decision-making in the areas of administration and instruction (Lynch, 2003). For example, by measuring and analyzing student outcomes with the assistance of an evaluator, teachers and administrators can quickly identify strengths and weaknesses of an instructional program, which may provide the necessary impetus for enacting curricular revisions, engaging in targeted professional development, and improving overall program design (Bernhardt, 2006).

Program evaluation employs similar research methods as those utilized in scholarly research; however, it also differs in a handful of important ways that make it a unique field with its own body of literature and standards governing its application (Stufflebeam & Coryn, 2014). For example, whereas scholarly research is aimed at testing theories and filling existing gaps with generalizable knowledge, evaluation addresses specific questions related to individual programs. Additionally, the questions being answered often originate from or are negotiated with the key stakeholders requesting their services (Newcomer, Hatry, & Wholey, 2015). In other words, program evaluation’s processes of practically oriented enquiry in situ are precisely the heuristic

for generating new knowledge, raising awareness, and transforming the educational, social, and economic circumstances of individuals and communities.

Evaluation in Foreign Language Education

According to Kiely (2009), the four main purposes for conducting an evaluation of a foreign language program include judging its worth, demonstrating its quality, increasing understanding, and improving effectiveness. Early applications of program evaluation in the field often focused on judging the worth of one instructional method compared to another. Examples include the Colorado Project (Scherer, 1969) and Pennsylvania Project (Smith, 1970), which attempted to empirically prove the merits of the audio-lingual and cognitive code methods, respectively. Next, recognizing the complex nature of language programs whose variables cannot easily be isolated for objective measurement, evaluation of foreign language programs in the 1970s attempted to capture the worth of an entire program by analyzing the products and processes of instruction (Kiely & Rea-Dickins, 2005). Following the introduction of the Program Evaluation Standards in 1981 (Yarbrough, 2011), responses in language education included increased attention on participatory evaluation, which sought to actively involve stakeholders such as teachers, students, and administrators in the evaluation process (Kiely & Rea-Dickins, 2009). Recently, research on effective foreign language teaching practices (Davin, Troyan, Donato, & Hellman 2011; Kearney, 2015; Troyan, Davin, & Donato, 2013) has more closely reflected Kiely's (2009) fourth purpose for evaluation (improving effectiveness) by engaging teachers in the process of understanding, developing, and improving their instructional practices.

To date there has been limited meaningful application of program evaluation in the field of foreign language education, particularly in the K–12 setting. This is unfortunate because, as Norris (2016) posited, this mode of inquiry may have strong potential to positively impact foreign language education practice, policy, and research. For example, recommendations by Davin, Rempert, and Hammerand (2014) included increasing consistency in the curricula, course offerings, methods of student placement, and time allotted for language instruction throughout an urban school district. Additionally, recommendations derived from a study by Kissau, Adams, and Algozzine (2015) extended beyond the local context and advocated for

expanding foreign language instruction as a required offering in middle schools nationwide. These findings reveal the inherent capacity for program evaluation to address current gaps in research and practice while also contributing to a greater understanding of what works in various contexts of language teaching and learning (Kiely & Rea-Dickins, 2009; Norris, 2016). Additionally, by objectively quantifying and describing the performance of a given program and providing recommendations for improvement, program evaluation provides a practical method by which to meaningfully respond to Sandrock's (2006) call for innovation in the field.

K–12 Foreign Language Outcomes

Language proficiency development is a highly variable and complex phenomenon that is dependent upon a host of individual, environmental, and program-level factors. Several early studies of K–12 students' oral proficiency indicated that the Intermediate Low level on the ACTFL Proficiency Guidelines (2012) is a reasonable goal for secondary learners to achieve upon program completion (Fall, Adair-Hauck, & Glisan, 2007; Glisan & Foltz, 1998; Huebner & Jensen, 1992). However, more recent studies that address students' ability level in all modalities as measured by standardized performance assessments problematize the feasibility of attaining Intermediate levels of proficiency, particularly when languages are only offered beginning at the high school level (Davin et al., 2014; Kissau et al., 2015). For example, in a recent report, Avant (2017) reviewed results from nearly 65,000 K–12 foreign language learners that took the STAMP4S during the 2015–16 school year. Findings revealed that Spanish students who completed four years of language study achieved an average rating approximately equivalent to Intermediate Mid for reading, and Intermediate Low for listening, speaking, and writing. Meanwhile, their French counterparts earned an average rating of Intermediate High for reading and Novice High for the remaining three skills. While this report reflects the inherent variability of outcomes across languages, levels, and modalities, it also provides the most comprehensive account of K–12 foreign language performance outcomes in recent history, and its outcomes are also leveled according to ACTFL Proficiency Guidelines (2012). For these reasons, this report will serve as the primary source for comparing outcomes in the current study. Regardless of the differences in research design, setting, and assessment type, the collection of studies reviewed

here generally indicate that the majority of students in K–12 foreign language programs will achieve at least Novice-High or Intermediate Low levels of proficiency.

Impact and Test Use

The impact of an assessment can be viewed as encompassing a breadth of consequences related to its utilization that are experienced by a variety of stakeholders in contexts that may extend beyond the classroom or school (Wall, 1997). Given the focus of this study, the discussion of impact will be limited to the way that educators modify their practices in response to student outcomes data. According to Bernhardt (2006), it is not enough to simply measure student outcomes or to look at the raw data that they produce; instead, this rich data needs to be interpreted and analyzed in meaningful ways that leads to the improvement of professional development, curricula, and language programs as a whole. This call to action is echoed by Carroll (2015), who stated: “The purpose of testing is always to render information to aid in making intelligent decisions about possible courses of action” (p. 43). Indeed, practitioners in our field have already begun the challenging work of interpreting the data gleaned from language assessments and applying that knowledge to make important decisions related to more than just classroom instruction (Purpura, 2016). Yet, there has been very little research that specifically addresses how educators interpret and respond to such data to improve their practices.

Knowledge Gap

The dearth of studies investigating language proficiency and performance outcomes in the K–12 setting, along with their limitations in generalizability, represents a clear and present need in the field of foreign language education research. Despite the well-established nature of the proficiency movement, and the lasting impact that documents such as the ACTFL Proficiency Guidelines (2012) and the World-Readiness Standards for Learning Languages (2015) have had on language instruction in the classroom (Liskin-Gasparro, 2003), there remains an acute lack of documentation surrounding the level of proficiency that learners are able to achieve following the completion of specific learning sequences (Malone, 2017). A second element that is either missing or underdeveloped in many of the studies of student outcomes (Davin et al., 2014; Fall et al., 2007; Glisan & Foltz, 1998), is an account of teachers’ agency in interpreting and utilizing assessment data as a means of professional development, or for facilitating curricular innovation

(Green, 2013; Rea-Dickins, 2004; Shrum & Glisan, 2016). Although teachers tend to lack adequate levels of assessment literacy (Popham, 2009), they nevertheless have a great deal of autonomy to enact changes to both what they teach and how they teach in response to various forms of assessment (Rea-Dickins, 2004). Local instances of utilization-focused program evaluation, such as the current study, are uniquely positioned to aid in the process of transforming assessment data into valuable knowledge, ultimately affecting decisions aimed at improving language teaching and learning (Kiely & Rea-Dickens, 2009).

The Present Study

Pursuant to the nature of the evaluation questions, which were negotiated with key district stakeholders to ensure the utility of the feedback, this study had two central foci. First, by measuring and analyzing students' performance outcomes, it aimed to hold the district accountable with respect to its stated goal for students to achieve Intermediate levels of proficiency upon program completion. Second, by surveying and interviewing teachers, the intent was to highlight current practices and future needs of the program, particularly with respect to the implementation of the AAPPL. These complimentary purposes for evaluation align with the categories of summative accountability and formative development as outlined by Rea-Dickins and Germaine (1992). Norris (2014, October) described the former as a regulatory mechanism and the latter as an educative process that leads to change and innovation, which is the primary goal of this evaluation.

Evaluation Questions

1. To what extent are students achieving Intermediate levels of proficiency in listening, reading, speaking, and writing as measured by the AAPPL?
2. What is the nature of world language teachers' reported instructional practices?
3. How do program stakeholders interpret and apply AAPPL outcomes data?
4. What are the professional development needs of world language teachers?

The Evaluator

Just as is the case in traditional scholarly research, objectivity on the part of the evaluator in program evaluations is of paramount importance to ensure that findings are accurate, justified, and lead to relevant and unbiased feedback for the program stakeholders (Yarbrough, 2011).

Arguably of equal importance is that evaluators are also experts or have adequate training in the content area of the program being evaluated (Norris, 2016). Thus, for the sake of transparency and to address these two areas of concern, the following description is offered of the evaluator's background and relationship to the current program being evaluated. The evaluator's foreign language teaching experience spanned eight years and multiple contexts, two of which are highlighted here. While teaching fourth grade in a French dual immersion program, this evaluator designed and delivered inquiry-based curricula that cultivated students' linguistic and academic development. Later, during four years as a high school French teacher in an urban K–12 district, he was actively involved in professional development on the applications of proficiency-oriented practices and collaboratively designed district-wide common assessments and curricula that aligned with ACTFL and state standards. Thus, in addition to more recently developing a skillset in applied research methods and program evaluation, he has adequate germane experience to ensure expert feedback in this context. Additionally, the evaluator was not affiliated with or compensated by the district in question. The opportunity to investigate the activity and outcomes of this K–12 world languages program was made possible because of the professional relationship that the evaluator developed in recent years with the district's World Languages Coordinator. However, this peripheral connection to the program did not have a bearing on the nature of the findings of the evaluation, which were offered as evidence for the objective recommendations for program improvement.

Methods

This mixed methods study was carried out through a pragmatic lens, meaning it involved the collection and synthesis of quantitative and qualitative data with the aim of understanding and depicting what works with respect to foreign language teaching, learning, and assessment in the local context (Moeller, 2016). Given the increasing complexity of purposes for and consequences of language assessments in the proficiency movement, mixed methods approaches have been increasingly applied when investigating practices surrounding language testing and assessment (Jang, Wagner, & Park, 2014). From the perspective of those conducting language program evaluations, there has been a similar call for the need to collect more than just assessment data in determining the actual performance of a given program (Norris, 2006, 2016).

Context

The study took place in a large urban school district with a total student population of more than 35,000, of which 59.5% received Free-and-Reduced-Lunch (FRL). Approximately 38% of the students in the district self-identified as African-American, with about 35% White students, 18% Hispanic, and the remainder of the students identified as Native American, Asian, or Other. In this district, world languages are taught predominantly in middle schools and high schools as an elective subject. During the 2017–18 school year, there were 92 world language teachers that provided instruction to approximately 34% of the entire K–12 student population. Meanwhile, approximately 77% of secondary students were enrolled in either a foreign or heritage language course.

Approximately ten years ago, an advisory committee for world languages outlined the mission, vision, and core beliefs of the program. As part of this initiative, a select group of teachers researched, negotiated, and articulated the instructional approaches, practices, and learning sequences that should be enacted throughout the district. The result was an emphasis on the importance of a standards-based, holistic approach to language acquisition that encouraged the use of Comprehensible Input¹ (CI) strategies in order to facilitate students' language proficiency development. In addition to district-wide curricula for Spanish and French, professional development (PD) and teacher coaching is consistently offered to support all world language teachers in the district. However, teachers' adherence to these materials, as well as their participation in the PD, is not explicitly required nor consistently monitored. Instead, they are afforded the autonomy to teach according to their own unique beliefs about foreign language teaching and learning.

Beginning in 2014, foreign language teachers were invited to administer the AAPPL to their students in Level Three and above, and heritage teachers could offer the assessment to their students at all levels of instruction. Based on their outcomes, students were able to receive either a World Language Proficiency Certificate for an Intermediate Low rating, or a Gold or Platinum Bilingual/Multilingual Seal for ratings of Intermediate High and Advanced Low, respectively.

¹ Comprehensible Input is defined here as an approach to language instruction that is predicated on Krashen's input hypothesis, which posits that learners will naturally acquire language when presented with language input that is comprehensible (Faltis, 2008).

Participants

The 47 teachers that completed the survey during Phase 1 of the study appeared to be largely representative of the district-wide population of 92 world language teachers in terms of the levels and languages of instruction (see Table 1). More than three quarters of participants (n=36) reported having earned a Masters degree, with the remaining 11 having completed a Bachelors degree. Their years of language teaching experience ranged from 2 to 36, with a mean of 12.3 years. As may be expected given the diversity of teachers' backgrounds and experiences, the languages that they teach, as well as the different populations of students that they serve, there were a wide variety of practices that this group of teachers reported using in their instruction.

Table 1. Survey Participants' Language and Level of Instruction

Language	Elementary	Middle school	High school	Total
Spanish	4	9	11	24
French		3	7	10
ASL			2	2
Arabic		1	1	2
Chinese			2	2
German			1	1
Japanese		1	1	1
Ojibwe		1		1
Spanish Heritage		2	1	3
Total	4	17	26	47

From the initial sample of teachers, 12 participants also agreed to be interviewed. Among this subset of teachers, eight taught at the high school level, three at the middle school level, and one at the elementary level. Meanwhile, there were five teachers of Spanish, five teachers of French, one teacher of Chinese, and one teacher of Spanish for Heritage Speakers. Additionally, four of the high school teachers administered the AAPPL previously and one was administering it for the first time. The district's World Languages Coordinator was also interviewed to provide an additional perspective on current practices with respect to teaching, curricula, and assessment.

In total, 20 foreign and heritage language teachers administered the AAPPL to 846 students, or 43% of those who were eligible to take the assessment. More specifically, 510 test takers were enrolled in Spanish as a foreign language, 146 were classified as Spanish Heritage learners, 107 were enrolled in French, 43 in Japanese, and 40 in Arabic. Of particular note, from among the 700 foreign language students assessed, 331 were ninth and tenth graders. In other words, nearly half of these learners likely began their language learning sequence while in middle school.

Research Design

The study was divided into three distinct but complementary phases. The study procedures are outlined in sequential order following a description of each instrument, which also served to demarcate each phase.

Survey Instrument

The teacher survey was adapted from multiple sources and revised to address the specific evaluation questions of this study (to view survey see: <https://www.iris-database.org/iris/app/home/detail?id=york%3a936111&ref=search>). The first of three main parts, adapted from Vyn, Wesely, and Neubauer (2019), addressed teachers' instructional practices and beliefs about language teaching and learning. Among other questions aimed at identifying their instructional approach, teachers were asked to identify the percentage of instructional time they devote to each of the four modalities, as well as the amount of target language they use in both beginner and upper-level classes. In the second part, which was derived from the Classroom Assessment Practices Questionnaire (CAPQ), designed by Cheng, Rogers, and Hu (2004), teachers were asked about their assessment-related activity, as well as how they utilize the resulting data to inform their practice. Finally, in part three teachers were asked to offer suggestions for future PD offerings they would like to receive.

Interview Protocol

The semi-structured interview protocol was designed by the evaluator and provided a framework for eliciting participants' description of their instructional and assessment-related practices (to view interview protocol see: <https://www.iris-database.org/iris/app/home/detail?id=york%3a936111&ref=search>). The protocol also provided an opportunity to confirm, clarify, or

elaborate on the predominantly quantitative information that teachers shared in the online survey. For example, teachers were asked to provide examples of the purposes for which they described utilizing assessment data. They were also asked more probing questions related to their interpretation of student score reports, as well as how they prepare students leading up to the assessment.

Assessment Instrument

The ACTFL Assessment of Performance toward Proficiency in Languages (AAPPL) is a standardized measure of language ability in the three communicative modes: interpersonal, presentational, and interpretive. It is comprised of four parts, Interpersonal Listening/Speaking, Presentational Writing, Interpretive Reading, and Interpretive Listening, and is generally administered online, taking approximately two hours to complete. There are two versions of the assessment, each of which is specifically designed to assess different ranges of proficiency levels. Form A is intended for use with learners whose proficiency level may range from Novice through Intermediate on the ACTFL Proficiency Guidelines (2012), and Form B is more appropriate for learners in the Intermediate to Advanced levels of proficiency. In each of the four components, learners complete contextualized tasks related to topics that are appropriate for the level of proficiency being tested.

Scoring for the Interpretive Reading and Listening portions of the assessment, which utilize multiple-choice answers, is completed automatically through computer-based software. Meanwhile, the Interpersonal Listening/Speaking and Presentational Writing portions are scored by independently certified AAPPL raters. Rather than being assigned an overall proficiency rating according to the ACTFL Proficiency Guidelines (2012), learners receive a numerical score within the particular proficiency band (i.e., N-4, which is approximately equivalent to Novice High) for each of the four components. An individual report is produced for each learner, indicating their score for each component and providing helpful strategies for improving their performance in each skill area.

Data Collection and Analysis

Phase 1

The study procedures began at a district-wide professional development (PD) day where eligible participants were invited to complete the online teacher survey. The survey was designed using Qualtrics and took approximately 10–20 minutes to complete. Initial analyses of participants' responses served as the basis for designing follow-up questions during Phase 2 interviews. Later, teachers' reported practices were illustrated through the use of simple descriptive frequencies, including charts and graphs. Additionally, teachers' open text responses to questions about their perceived PD needs were tabulated to identify the most frequently cited topics.

Phase 2

The second phase of the study involved scheduling and carrying out the semi-structured interviews, which were conducted either over the phone or via Zoom, a web-based video-conferencing software. During the interviews, which lasted approximately 45 minutes to an hour, the evaluator followed interview guidelines as outlined by Weiss (1994). The analysis of interview data was oriented around identifying key themes and quotes that characterize teachers' instructional and assessment-related practices. To that end, all interviews were first transcribed in their entirety by the evaluator, then analyzed using a grounded theory approach to text analysis, as characterized by Corbin and Strauss (1990). Where possible, connections were made across interviews to identify and describe the consistency or prevalence of various themes.

Phase 3

The AAPPL was voluntarily administered by 20 teachers to 846 eligible upper-level foreign and heritage language students during their regularly scheduled class periods toward the end of the 2017–18 academic year. Months later, following the receipt of de-identified AAPPL outcomes data from the district, the evaluator first removed invalid data and ensured that student scores would be compared to their appropriate peer groups by language and level. During these procedures, the outcomes for foreign language (n=700) and heritage language learners (n=146) were split into separate files for independent analyses, with attention being paid here solely to the performance of foreign language learners. Additionally, Arabic and Japanese outcomes were excluded from analysis due to their small sample size. Next, students' individual raw scores for

each of the four assessment components were converted into a numeric scale (i.e., N-4 = 4, I-2 = 6, and A = 10), and composite scores were calculated based on the average of all four numeric scores. Finally, descriptive statistics including charts and graphs were produced to illustrate district-wide average scores in all four modalities, as well as the composite totals, for each language and level tested.

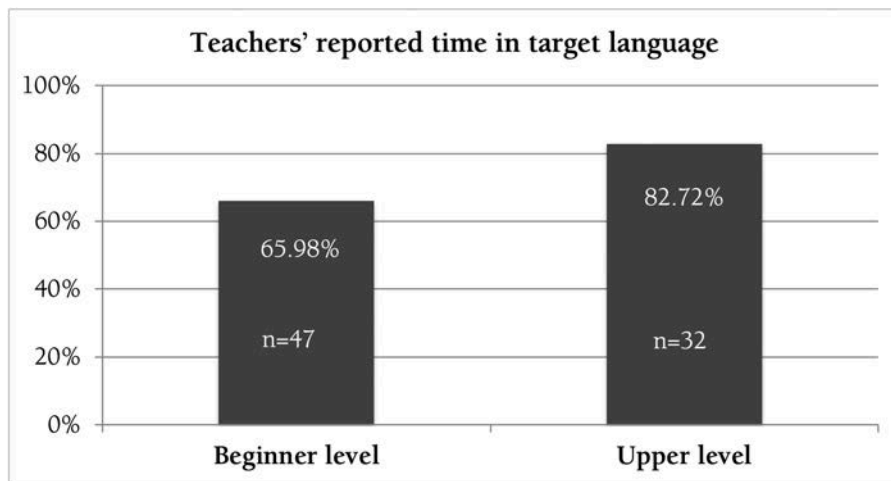
Findings and Discussion

Although the presentation and discussion of findings herein will necessarily make connections to the literature, the primary purpose is to provide a logical framework for the evidence in order to justify recommendations provided to the local program stakeholders.

Instructional and Assessment-Related Practices

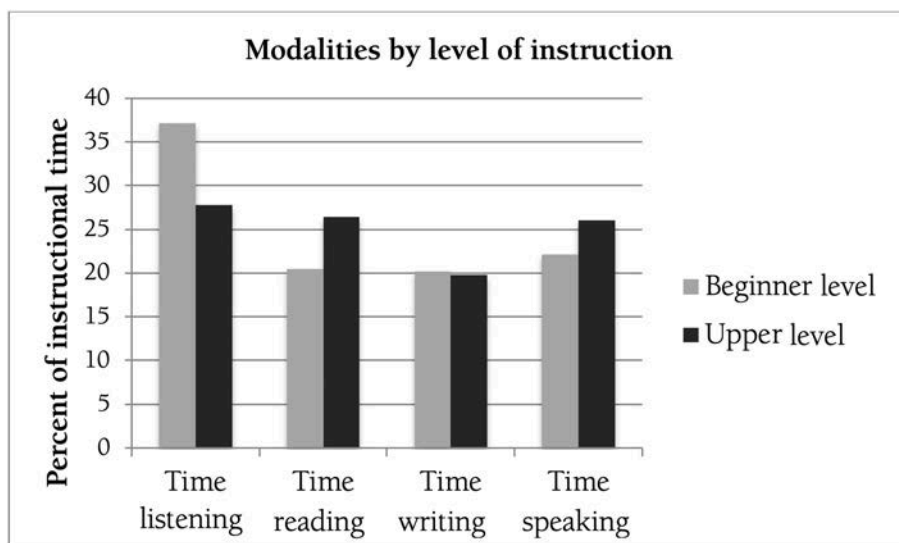
Despite the wide variety of practices teachers reported enacting to meet the individual needs and interests of their students, there were two distinct trends that emerged in beginner and upper-level classes. First, teachers' estimated use of the target language was significantly lower in the beginner-levels than in upper-levels (see Figure 1). This disproportionate use of the target language, which appeared to be consistent across all languages, should be considered in light of ACTFL's position statement that advocated for "language educators and their students use the target language as exclusively as possible (90% plus) at all levels of instruction" (ACTFL, 2010). Additionally, Vyn et al. (2019) also found that the relationship between teachers' reported time in the target language and student outcomes, while positively correlated with student outcomes across nearly all levels, was most pronounced during the earlier levels of instruction. While teachers' average reported time in the target language in the upper levels is approaching the benchmark goal of 90%, there remains much room for growth at the beginning levels of instruction.

Figure 1. Time in Target Language



The importance of increasing teachers' time in the target language is further heightened in light of the percentage of instructional time they devote to each modality. As illustrated in Figure 2, teachers reported allocating more time to listening compared to the remaining three skills in beginner-level classes. Thus, it is essential for teachers to ensure a high degree of target language use to provide learners with adequate exposure to the language for acquisition to occur. Transitioning to the upper-level courses, the more balanced distribution of time devoted to each modality may be attributed in part to the fact that the AAPPL itself tests each modality separately. Regardless of the reason for the current distribution of instructional time, this serves as a framework from which to better understand the ways that teachers reported integrating the AAPPL data into their planning and instruction.

Figure 2. Use of Modalities



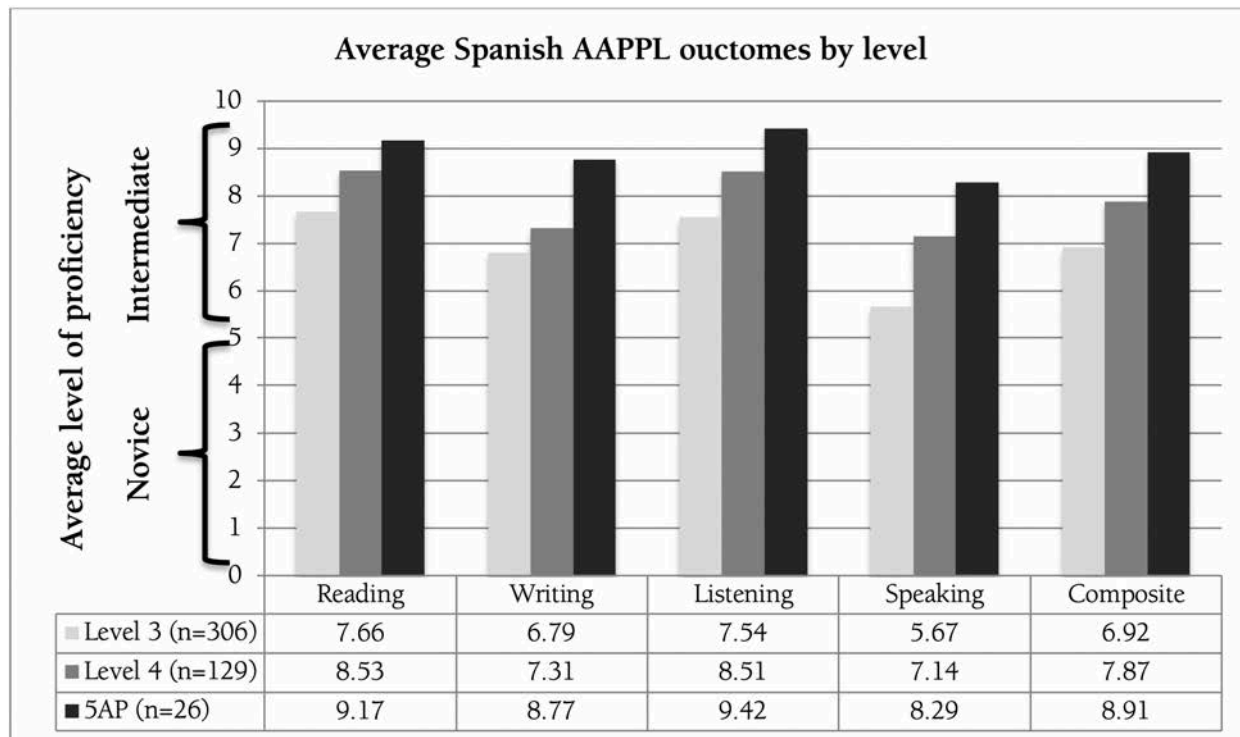
Student Outcomes

Spanish

Average performance outcomes for students of Spanish as a foreign language ranged from Intermediate Mid in Level Three to Intermediate High in Level Five/Advanced Placement (see Figure 3). Therefore, not only is the Spanish program exceeding the district's benchmark proficiency goal, but it is also consistently measuring at or above national average outcomes for each individual modality (Avant, 2017). Particularly noteworthy among the eight Spanish teachers whose students' results are illustrated here is the extent to which they reported integrating proficiency-oriented and input-focused methods in their instruction. All four Spanish teachers from among this group that responded to the survey indicated that they not only use CI strategies, but they also use the target language at least 90% of the time. Thus, the nature of their instructional practices may partially explain why students' receptive skills (reading and listening) were both consistently higher at each level than were their production skills (writing and speaking). Although not by a large margin, Spanish students were found to be weakest at speaking across all levels. This finding is consistent with national averages as reported by Avant (2017) wherein student outcomes in speaking were among the weakest of the modalities and demonstrated the least amount of variability over the five levels of instruction. With respect to the varying rate of development in each modality, it should be unsurprising that speaking is later to develop among learners in this sample. Nevertheless, by Level Four they were performing on average at the Intermediate Mid level and approaching the Intermediate High level, which

surpasses the average levels of oral proficiency found in a similar population of students in Glisan and Foltz (1998) and was approximately on par with findings in Fall et. al (2007).

Figure 3. Spanish Outcomes by Level



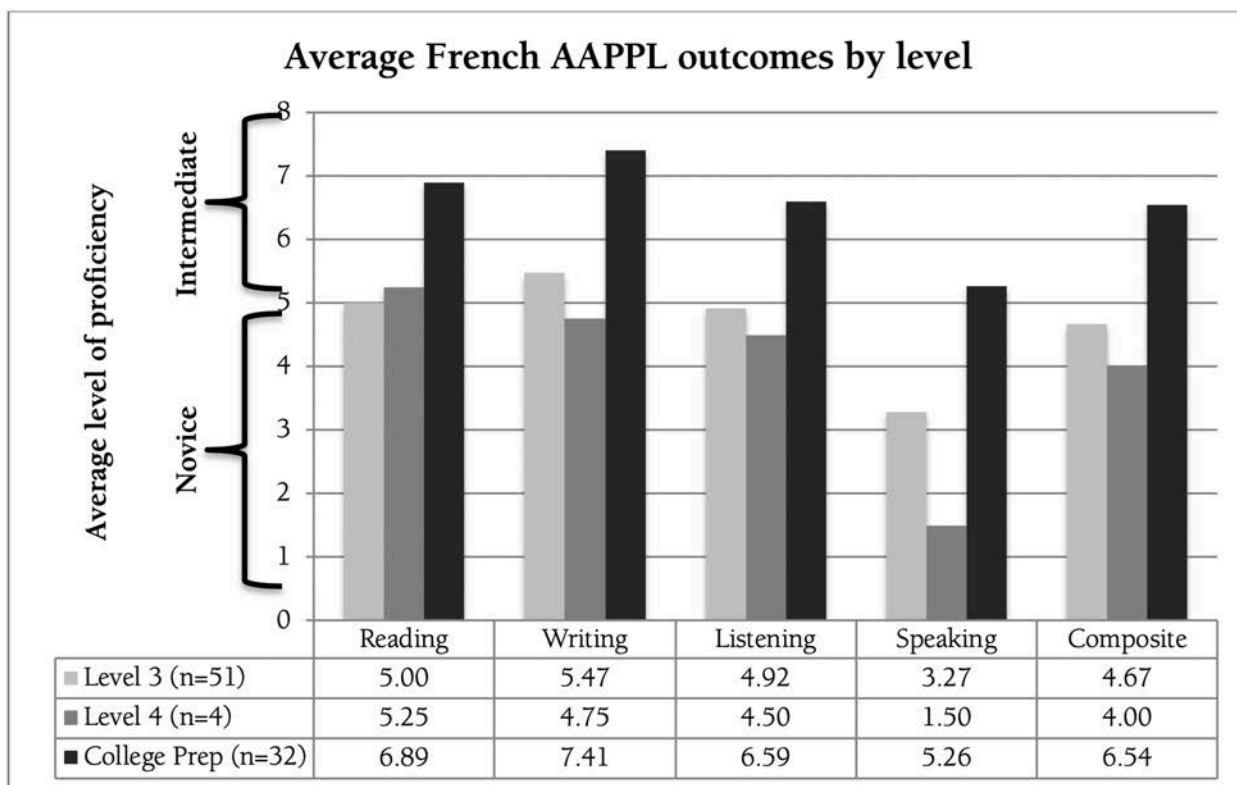
KEY: The numbers 1–10 represent the following AAPPL scores: 1=N-1, 2=N-2, 3=N-3, 4=N-4, 5=I-1, 6=I-2, 7=I-3, 8=I-4, 9=I-5, 10=A. The numbers for each skill indicate the average achievement for that group.

French

Although the French program underperformed with respect to the district’s goal for students to achieve Intermediate levels of proficiency according to the ACTFL Proficiency Guidelines (2012), AAPPL scores were approximately on par with national averages of K–12 French performance outcomes (see Figure 4). Composite averages for Level Three and Four students are both in the Novice High range, with outcomes for the modality of speaking being even further off the mark for French students as compared to their Spanish counterparts. It should be noted that students enrolled in College Prep who earned an average composite rating of Intermediate Mid belong to a more selective and rigorous course, which requires at least three years of prior language enrollment and enables students to concurrently earn college credit. Their average performance exceeded national averages for French learners with similar years of experience in all skill areas with the exception of reading (Avant, 2017). However, despite their comparatively strong

performance, their composite score still falls short of average performance outcomes attained by Level three Spanish students. Regarding the disparate nature of the outcomes between French and Spanish students, there are several contextual factors that may be at play. First, the sample size of French students (n=87) and teachers (n=3) is significantly smaller than that of Spanish, for which there were 461 students across eight teachers. Additionally, the FRL populations among the schools represented in the French results are all higher than the district average. Since the FRL population has been found to negatively correlate with outcomes in this and a previous study of foreign language outcomes (Vyn et al., 2019), this may partially explain the observed differences in outcomes. Finally, the instructional methods and average amount of time in target language reported by the two French teachers whose students comprise all Level Three and Four outcomes stand in contrast to the predominantly proficiency-oriented practices reported by their Spanish teaching counterparts.

Figure 4. French Outcomes by Level



KEY: The numbers 1–8 represent the following AAPPL scores: 1=N-1, 2=N-2, 3=N-3, 4=N-4, 5=I-1, 6=I-2, 7=I-3, 8=I-4. The numbers for each skill indicate the average achievement for that group.

This variability in outcomes across languages, levels, and modalities underscores the notion that language proficiency development is a highly complex phenomenon that is

dependent upon a host of individual, environmental, and program-level factors (Davin et al., 2014; Green, 2013). Although there is not enough evidence in the data gathered in this study to make empirical claims regarding the relationship between teachers' reported practices and student outcomes, it would appear based on the salient differences in performance between Spanish and French students within this district that it may be necessary to further investigate the role of target language use and the high-leverage practice of Facilitating Target Language Comprehensibility (Glisan & Donato, 2017).

Responding to Outcomes Data

Planning and Instruction

All three interviewees that previously administered the AAPPL reported adjusting the amount of instructional time devoted to the modalities in an attempt to remedy negative trends in students' performance. For example, when talking about differences in student outcomes, Everett (HS French) described, "Whatever the class comes out weakest in, I have to focus at, because I haven't been hitting that well enough for the students." Meanwhile, Krista (Heritage Spanish) responded similarly to trends in the data by evaluating how her judicious use of the modalities might better support their language learning progress. She stated, "So that's why, as I've worked here for four years, it's become more written based, because I see on the test—that's the area that they primarily struggle in." Finally, Teresa (HS Spanish) echoed a similar sentiment, although she described using students' AAPPL data to affirm or challenge her own intuition with respect to their individual areas of weakness. In sum, teachers focused a great deal of attention on calibrating the amount of instructional time allocated to engaging students in each skill (listening, reading, speaking, and writing) based on average student performance on each component of the test. This intentional response to student outcomes data was similar to the consciousness-raising effect for teachers that were involved in the design and administration of an oral assessment in Fall et al. (2007). In both instances, the administration of the assessment served to reinforce teachers' enactment of proficiency-oriented approaches to instruction and to more deliberately integrate the standards into their practice.

Interactions with Students

Everett, Krista, and Teresa each took various approaches to debriefing AAPPL results with students, and for different reasons, too. For example, at the most basic level, Teresa articulated needing to interpret the results for her students so that they more fully understood their ability level. In reference to the score report, she said, “that little print out sheet to them, initially they were like, ‘What does this mean, a one?’” Over time, Teresa recounted not needing to spend as much time facilitating the interpretation and analysis of results, both as the reports became more decipherable, but also as the notion of proficiency and the modalities became common vernacular for her students. For Krista, the discussion of outcomes with her heritage students centered more on communicating whether or not they earned the Seal of Biliteracy. Then, if they fell short, the strategies for improvement included in the AAPPL score report required further interpretation on Krista’s part because they are oriented toward foreign language students. To that effect, she stated, “I am interpreting their results as native speakers of a test that is not designed for them, right?” Krista also reported meeting individually with her students to discuss their perceptions of the assessment and the factors that may have affected their performance. From there, they discussed key strategies for continued improvement before possibly retaking the AAPPL the following year. Everett, on the other hand, discussed the outcomes more generally as a whole class. He reported emphasizing to his students the importance of putting forth their best effort on these and other high-stakes tests. Alluding to the power of such tests and the gravity of the decisions made based on the results, he attempted to instill in them a sense of urgency to do their best.

Program Advocacy

One final use for the AAPPL outcomes data that was observed among stakeholders in this district, and which touches on the potentially high-stakes nature of this assessment, was for program advocacy. For example, Teresa described using her students’ AAPPL results to demonstrate to her administrator how well her classes were going. To that end, she stated, “part of it’s PR (public relations)—look at, kids are learning, this is a yardstick of what they’re doing, and this is how they’re achieving.” Thus, she was able to provide evidence of the quality of her teaching through student outcomes data. On a much larger scale, the World Language Coordinator described presenting the district-wide outcomes on the AAPPL to the school board

and other high-ranking school officials. In particular, she highlighted the number of students that earned the World Language Certificate and Bilingual Seal across all languages, thereby promoting the value and importance of world language instruction more broadly. Each of these examples are also a manifestation of Bernhardt's (2006) call for such language assessment data not only to be leveraged as a public relations tool, but also to make informed programmatic and instructional decisions.

In summary, although the extent to which teachers reported using the AAPPL outcomes data varied greatly, the primary purposes were to inform changes to planning and instruction, to facilitate students' awareness of and goal setting for language proficiency development, and as means to demonstrate the effectiveness of instruction and promote the importance of language learning more broadly.

Professional Development Needs

Teachers frequently expressed interest in (a) understanding the content and outcomes of the AAPPL and how to interpret and apply that knowledge meaningfully in practice, and (b) collaborating across language levels to design vertically and horizontally articulated curricula. These two areas of need were very much interrelated. For example, many of the participants (n=9) who described wanting increased opportunities for collaboration to more closely align curricula and instruction also described wanting to become more familiar with the content of the AAPPL. Even teachers whose students do not take the AAPPL described that having a better knowledge of what is expected of students by the end of high school would help with aligning curricula and making informed decisions related to their planning and instruction. As one such teacher described, "The more we know as middle school teachers about these tests, the more we can prepare our students early on." Meanwhile, all four interviewees that implemented the AAPPL in 2017–18 indicated wanting to learn how to interpret the score reports and leverage that knowledge to better align their instruction to meet students' needs. However, these teachers also reported obstacles such as time, resources, and lack of familiarity with the reports that currently prevent them from realizing this goal. This finding is perhaps unsurprising given the widespread lack of assessment literacy reported among language educators (Popham, 2009). Regardless, the common thread among these self-reported areas of interest for future professional

development is a desire among teachers to improve the effectiveness of their instruction for the benefit of student learning and achievement. In other words, by actively seeking to align their instruction with the construct of the AAPPL, they are demonstrating their agency in response to the assessment (Rea-Dickins, 2004).

Recommendations

To ensure the continued success of the district's world languages program, the following recommendations are offered with the aim of facilitating improvements to the support and professional development provided to all world languages teachers.

Instructional and Assessment-Related Practices

- Given the current level of target language use across the district, teachers should be supported with professional development that focuses on practical strategies for meeting ACTFL's recommended benchmark of 90% or more.
- Provide increased opportunities for collaboration among and across languages and instructional levels to facilitate vertical and horizontal articulation of curricula and instructional practices, and to enable open sharing of ideas and resources between peers.

AAPPL Implementation and Outcomes

- Continue to expand the implementation of the AAPPL to provide students increased opportunities for recognition of their functional language ability and to facilitate teachers' reflective practices and ongoing curricular innovation.
- Offer PD for teachers that addresses how to meaningfully align planning and instruction with the AAPPL, and how to interpret and make informed decisions based on the available data.

Stakeholder Utilization of Recommendations

It is worth noting that since the dissemination of these results has already facilitated curricular innovation in the local context. After sharing a technical evaluation report with all district stakeholders, the evaluator was invited back to coordinate a PD workshop for world language teachers that specifically addressed two of the above recommendations. In addition to presenting an overview of the AAPPL's construct, content, and format for all world language teachers, the evaluator also guided teachers that administered the assessment through a protocol that focused on how to analyze and interpret student outcomes, as well as leverage that data to make evidence-based adjustments in their planning and instruction.

Study Limitations

The findings and recommendations are based on self-reported survey responses from slightly more than half of the district's world language teachers. Additionally, only three teachers that previously implemented the AAPPL agreed to be interviewed, and their practices may not be representative of the experiences of the larger group of twenty teachers that have administered the assessment. To mitigate such limitations in subsequent evaluations, evaluators should consider conducting longitudinal investigations, beginning with the collection of baseline data prior to the program's implementation of a given assessment. Additionally, by incorporating classroom observations at strategic intervals, researchers could triangulate teachers' self-reported practices and provide a more comprehensive description of changes to instruction over time.

Conclusion

The current study documented the instructional practices and student outcomes in the context of an urban K–12 world languages program. Framed by a practical approach to research and building on the literature surrounding student outcomes in this population, this study evaluated the program's ability to meet its stated objective of facilitating students' development of Intermediate levels of proficiency. In addition to contributing to the limited body of research in the field surrounding K–12 language performance outcomes, the study also demonstrated the potential for local instances of program evaluation to foster innovation in foreign language teaching and learning (Kiely & Rea-Dickens, 2009). Finally, it demonstrated the need for continued research of its kind to develop a more comprehensive understanding of the interaction between language assessment and teaching and learning in K–12 programs.

References

- American Council on the Teaching of Foreign Languages (ACTFL). (2010). *Use of the target language in the classroom*. Retrieved from <https://www.actfl.org/news/position-statements/use-the-target-language-the-classroom>
- American Council on the Teaching of Foreign Languages (ACTFL). (2012). *ACTFL proficiency guidelines 2012*. Alexandria, VA: Author. Retrieved from http://www.actfl.org/sites/default/files/pdfs/ACTFLProficiencyGuidelines2012_FINAL.pdf
- Avant (2017). *Avant STAMP results—national averages 2016–2017*. Retrieved from <https://d3itqtxdxl1nz0.cloudfront.net/pdfs/STAMP-National-Averages-Update-2017.pdf>
- Bernhardt, E. B. (2006). Student learning outcomes as professional development and public relations. *Modern Language Journal*, 90(4), 588–590.
- Carroll, J. B. (2015). Fundamental considerations in language testing. In A. J. Kunnan (Ed.), *Language Testing and Assessment* (pp. 43–51). New York, NY: Routledge.
- Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructor's classroom assessment practices: Purposes, methods and procedures. *Language Testing*, 21(3), 360–389.
- Corbin, J., & Strauss, M. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, 13(1), 3–21.
- Cox, T. L., Malone, M. E., & Winke, P. (2018). Future directions in assessment: Influences of standards and implications for language learning. *Foreign Language Annals*, 51(1), 104–115.
- Davin, K., Troyan, F. J., Donato, R., & Hellman, A. (2011). Research on the integrated performance assessment in an early foreign language learning program. *Foreign Language Annals*, 44(4), 605–625.
- Davin, K., Rempert, T., & Hammerand, A. (2014). Converting data to knowledge: One district's experience using large-scale proficiency assessment. *Foreign Language Annals*, 47(2), 241–260.
- Donato, R., Tucker, G. R., Wudthayagorn, J., & Igarashi, K. (2010). Converging evidence: Attitudes, achievements, and instruction in the later years of FLES. *Foreign Language Annals*, 33(4), 377–393.
- ESSA (2015). Every Student Succeeds Act of 2015, Pub. L. No. 114–95 § 114 stat.
- Fall, T., Adair-Hauck, B., & Glisan, E. (2007). Assessing students' oral proficiency: A case for online testing. *Foreign Language Annals*, 40(3), 377–406.
- Faltis, C. (2008). Comprehensible input. In J. M. González (Ed.), *Encyclopedia of Bilingual Education* (pp. 163–168). Thousand Oaks, CA: SAGE Publications, Inc.
- Glisan, E. W., & Foltz, D. A. (1998). Assessing students' oral proficiency in an outcome-based curriculum: Student performance and teacher intuitions. *Modern Language Journal*, 82(1), 1–18.

- Glisan, E. W., & Donato, R. (2017). *Enacting the work of language instruction: High-leverage teaching practices*. Alexandria, VA: The American Council on the Teaching of Foreign Languages (ACTFL).
- Green, A. (2013). Washback in language assessment. *International Journal of English Studies*, 13(2), 39–51.
- Huebner, T., & Jensen, A. (1992). A study of foreign language proficiency-based testing in secondary schools. *Foreign Language Annals*, 25(2), 105–115.
- Jang, E., Wagner, M., & Park, G. (2014). Mixed methods research in language testing and assessment. *Annual Review of Applied Linguistics*, 34, 123–153.
- Kearney, E. (2015). A high-leverage language teaching practice: Leading an open-ended group discussion. *Foreign Language Annals*, 48(1), 100–123.
- Kiely, R. (2009). Small answers to the big question: Learning from language programme evaluation. *Language Teaching Research*, 13(1), 99–116.
- Kiely, R., & Rea-Dickins, P. (2005). *Program evaluation in language education*. New York, NY: Palgrave Macmillan.
- Kiely, R., & Rea-Dickins, P. (2009). Evaluation and learning in language programmes. In K. Knapp & B. Seidlhofer (Eds.), *Handbook of foreign language communication and learning* (pp. 663–694). New York, NY: Mouton de Gruyter.
- Kissau, S., Adams, M. J., & Algozzine, B. (2015). Middle school foreign language instruction: A missed opportunity? *Foreign Language Annals*, 48(2), 284–303.
- Liskin-Gasparro, J. E. (2003). The ACTFL proficiency guidelines and the oral proficiency interview: A brief history and analysis of their survival. *Foreign Language Annals*, 36(4), 483–490.
- Lynch, B. K. (2003). *Language assessment and programme evaluation*. Edinburgh, Scotland: Edinburgh University Press.
- Malone, M. (2017, July 11). *Challenges and opportunities world in language research*. ACTFL CARD perspective. ACTFL Research SIG.
- Malone, M., & Sandroock, P. (2016). Assessment issues in world languages. In H. Braun (Ed.), *Meeting the challenge to measurement in an era of accountability* (pp. 217–235). New York, NY: Routledge.
- Moeller, A. (2016). The confluence of language assessment and mixed methods. In A. Moeller, J. W. Creswell, & N. Saville (Eds.), *Second language assessment and mixed methods research* (pp. 3–16) Cambridge, England: Cambridge University Press.
- National Standards Collaborative Board. (2015). *World-readiness standards for learning languages* (4th ed.). Alexandria, VA: Author.
- Newcomer, K. E., Hatry, H. P., & Wholey, J. S. (2015). *Handbook of practical program evaluation*. Hoboken, NJ: Wiley.
- No Child Left Behind Act of 2001, P.L. 107–110, 20 U.S.C. § 6319, (2002).

- Norris, J. M. (2006). The why (and how) of assessing student learning outcomes in college foreign language programs. *Modern Language Journal*, 90(4), 576–583.
- Norris, J. M. (2014, October). *Using evaluation for innovation and improvement in language and humanities education*. Paper presented at the Language Program Evaluation Symposium, Harvard University, Cambridge, Massachusetts.
- Norris, J. M. (2016). Language program evaluation. *Modern Language Journal*, 100(S1), 169–189.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48, 4–11.
- Purpura, J. E. (2016). Second and foreign language assessment. *Modern Language Journal*, 100(S1), 190–208.
- Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing*, 21(3), 249–258.
- Rea-Dickins, P., & Germaine, K. (1992). *Evaluation*. Oxford, England: Oxford University Press.
- Ricardo-Osorio, J. G. (2008). A study of foreign language learning outcomes assessment in U.S. undergraduate education. *Foreign Language Annals*, 41(4), 590–610.
- Sandrock, P. (2006). Meeting the challenge of tomorrow. *Foreign Language Annals*, 39(3), 357–357.
- Scherer, G. A. C. (1969). *A psycholinguistic experiment in foreign-language teaching*. New York, NY: McGraw-Hill.
- Shrum, J. L., & Glisan, E. W. (2016). *Teacher's handbook: Contextualized language instruction* (5th ed.). Boston, MA: Cengage.
- Smith, P. D. (1970). *A comparison of the cognitive and audio-lingual approaches to foreign language instruction: The Pennsylvania foreign language project*. Philadelphia, PA: The Center for Curriculum Development, Inc.
- Stufflebeam, D. L., & Coryn, C. L. S. (2014). *Evaluation theory, models, and applications* (2nd ed.). Hoboken, NJ: Wiley.
- Troyan, F. J., Davin, K. J., & Donato, R. (2013). Exploring a practice-based approach to foreign language teacher preparation: A work in progress. *The Canadian Modern Language Review/La revue canadienne des langues vivantes*, 69(2), 154–180.
- Vyn, R., Wesely, P., & Neubauer, D. (2019). Exploring the effects of foreign language instructional practices on student proficiency development. *Foreign Language Annals*, 52(1), 45–65.
- Wall, D. (1997). Test impact and washback. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language education* (Vol. 7, pp. 291–302). Dordrecht, Netherlands: Kluwer.
- Weiss, R. (1994). *Learning from strangers: The art and method of qualitative interview studies*. New York, NY: The Free Press.

Yarbrough, D., B. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: SAGE.

Contributors

Emilia Illana Mahiques is a Ph.D. candidate in Second Language Acquisition and Spanish Graduate Teaching Assistant at the University of Iowa, with educational background in education, pedagogy, curriculum design, and Spanish studies. Her research interests include online peer reviewing, digital narratives, second language writing development, Computer Media Communication in second language writing, and technology in language teaching and learning.

Anna Evans is a Ph.D. candidate in Second Language Acquisition at the University of Iowa. Her research interests lie in intersections between second language acquisition and second language assessment in a classroom context. Her dissertation project focuses on the development and assessment of writing skills and presents an in-depth analysis of French learners' performances on a source-based writing task.

Ari Natarina received her Ph.D. from the Department of Linguistics at the University of Iowa in 2018. Her research interests are language documentation, Austronesian syntax, sentence processing, and second language acquisition.

Runqing Qi is a Ph.D. student in Second Language Acquisition at the University of Iowa. She has extensive Chinese L2 teaching experience for both adult and child learners. Her research focuses on how to use technology tools effectively in L2 classrooms and Chinese L2 reading instruction. Her current project explores the use of authentic materials in Chinese L2 reading instruction.

Janet Stewart is a graduate student at the University of Minnesota-Twin Cities, where she is pursuing a master's degree in TESOL. She has a bachelor's degree in Linguistics and Russian language and literature from the University of Wisconsin—Madison. Her current research interests include cross-linguistic influence, intercultural communication, and foreign language assessment.

Reuben Vyn is a Ph.D. candidate in Foreign Language and ESL Education at the University of Iowa. His research interests include the relationship between foreign language instructional practices and learners' proficiency development, the use of language performance assessment in measuring and facilitating curricular innovation, and utilization-focused language program evaluation.



This CARLA working paper is available from the CARLA Website.

carla.umn.edu/resources/working-papers/

CARLA | Center for Advanced
Research on Language
Acquisition

UNIVERSITY OF MINNESOTA