EFFECTIVENESS OF SEMESTER-LONG VIDEO-SYNCHRONOUS CONVERSATION PRACTICE

by

Ayaka Matsuo

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Arts

School of Languages and Cultures

West Lafayette, Indiana

May 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Atsushi Fukada, Chair
Department of Art and Design

Dr. M. Mariko Wei
Department of Art and Design

Dr. Kazumi Hatasa
Department of Curriculum and Instruction

Approved by:
Dr. Jennifer M. William

ACKNOWLEDGMENTS

I thank my major professor, Prof. Atsushi Fukada, from the bottom of my heart for his enthusiastic and sincere support. Also, without previous advice from Prof. Wei and Prof. Hatasa, my thesis would not have completed. I am very grateful for them, too. Last but not least, I would like to show my gratitude to all the people who supported me in this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This study empirically investigated the effect of semester-long video-synchronous conversation practice on L2 Japanese learners' improvement of oral proficiency. The participants were 31 intermediate JFL university students at a U.S. university, and participated in the experiment as a part of their coursework. This study deployed a pre-post research design, and analyzed students' improvement of oral proficiency quantitatively.

Based on the previous finding that learners may benefit more from conversations with a non-native partner than with a native one (Varonis and Gass, 1985), the present research examined whether or not there are differences on the effectiveness of online conversation practice depending on interlocutor type (i.e., peer or native as a conversation partner). Accordingly, the experiment was conducted with a control group (Kanji Group), and two experimental groups (Peer Group and Native Group) in order to examine 1) the effectiveness of weekly video-synchronous conversation practice on learners' improvement of oral proficiency, and 2) the differences on the effectiveness of the practice by the interlocutor type.

The results found that all the three groups significantly improved their oral proficiency, and there were no differences detected between the two experimental groups. Since the control group also made significant gains, evidence for the effect of the conversation practice was not found. With regard to the second question, assuming that a large portion of the improvement of the two experimental groups was accounted for by the effect of the online practice, whether or not the conversation partner was a peer or a native did not make a difference.

Along with the quantitative analysis of students' improvement, this study reports students' perceptions toward the practice and their beliefs about conversation partners' native speaker status. Regardless of the interlocutor type, they showed overall positive reactions to the practice. As for their beliefs about conversation partners' native speaker status, strong preferences for native speaker as a conversation partner were found in the participants in all three groups.

CHAPTER 1. INTRODUCTION

This thesis will investigate the effectiveness of weekly video-synchronous conversation practice with either a peer or a native speaker partner as a part of coursework. Since the motivation for the present research is to seek out effective, practical, and educationally sound ways of conducting speaking practice outside the classroom, students' perceptions toward the practice will also be examined. In addition, in order to maximize students' learning from the practice, students' beliefs about conversation partners' native speaker status will be explored.

This chapter will present backgrounds and motivations of the present research. Specifically, the status of oral proficiency as a priority area, and relationship between students' beliefs and learning effect will be discussed below. Then, research questions will be formulated at the end.

Oral Proficiency

The purpose of the present research is to investigate effective, practical, and educationally sound ways to improve students' oral proficiency. The present author, as a graduate instructor of Japanese in the U.S., became motivated to investigate the topic by both teachers' and learners' high interests in speaking skill in language learning and teaching.

The American Council on the Teaching Foreign Languages (ACTFL) considers oral proficiency to be crucial in foreign language education in the United States. In their Standards for Foreign Language Learning (1996), they stated their philosophy as follows:

> The United States must educate students who are linguistically and culturally equipped to communicate successfully in a pluralistic American society and abroad. This imperative envisions a future in which ALL students will develop and maintain proficiency in English and at least one other language, modern or classical. (p.2)

Although the statement is not limited to oral skills, it is lucid that foreign language education in the U.S. attaches importance to oral proficiency, considering the fact that foreign language teachers are expected to speak the target language 90% of the time in the classroom (Kissau, 2014), and 21 states require foreign language teacher candidates to be orally proficient in the target language (Chambless, 2012) as follows; the minimum level required for French, German, Hebrew, Italian,

Portuguese, Russian, and Spanish is Advanced Low as defined by the Oral Proficiency Interview (OPI) guidelines, and for Arabic, Chinese, Japanese, and Korean, the minimum requirement is Intermediate High. (ACTFL, n.d.).

Motivated by the ACTFL proficiency guidelines, language teachers tend to attach weight to speaking in classroom teaching. Fukada and Wei (2012) surveyed JFL and JSL teachers from all levels (from beginner to advanced), and revealed that teachers from all levels consider teaching of speaking and listening as more important than that of reading and writing. Further, comparing speaking and listening, they place emphasis on teaching of speaking over that of listening.

Oral proficiency is recognized as an important component in foreign language education not only by educators but also by learners. Miller (2019) investigated goals of 62 L2 Spanish learners in the U.S., and reported that many of them put the highest priority on speaking (74.3%) followed by grammar (37.1%), listening comprehension (27.4%), and fluency (24.2%).

As shown above, becoming orally proficient in the target language is a crucial goal for both teachers and learners. Pertinent to the improvement of oral proficiency, various scholars formulated hypotheses as to how language acquisition occurs. For example, Krashen (1985) suggested the Input Hypothesis, which hypothesizes that language acquisition occurs through comprehensible input. Swain (1985) proposed the Output Hypothesis, in which she promotes that learners output the target language for successful language acquisition. Long (2015) advocated the Interaction Hypothesis, in which he explains that interactions can help language acquisition.

Following the hypotheses, foreign language teachers today strive to create as many opportunities as possible for students both inside and outside a classroom where they can interact with others using the target language. However, in-class interaction is extremely limited in terms of format and time for each student to speak. Discussing in-class speaking practice, Fukada and Wei (2012) estimated that the amount of time for an individual student to be involved in communicative speaking practice would be less than 10 minutes in a 60-minutes class when 25 minutes are devoted to oral practice.

In order to improve this situation, teachers and researchers started to seek out extra interaction opportunities outside a classroom. For instance, Kato et al. (2016), and Saito and Akiyama (2017) investigated the effectiveness of language exchange projects in improving students' oral proficiency that connected university students learning each other's first languages in the U.S. and in Japan on online video communication tools (e.g., Skype and Google Hangout).

Terhune (2016) reported on a possibility of extended interaction opportunities for students by integrating private online lessons with well-trained EFL teachers in the Philippines into ongoing coursework. Thus, these recent innovations offer a greater number of interaction opportunities for language learning than before.

Demands for improving oral proficiency are high in language learning and teaching, and researchers are on the constant lookout for opportunities that new technologies might present. Among these new innovations, video-synchronous communication tools are highly attractive innovations for language teachers because they can remotely create an interactional environment similar to an ordinary face-to-face one. As presented earlier, researchers actively search for effective ways of using such new tools. However, to date, empirical research that examines effectiveness of activity designs utilizing video-synchronous communication tools is not sufficient yet.

Students' Beliefs and Learning Effect

Just as we have beliefs about many things in our life, language learners should have beliefs specifically about language learning. It should be interesting just to find out what their beliefs are about language learning, but if beliefs influence their learning achievement either positively or negatively, it becomes essential for teachers to study students' beliefs about language learning.

According to Wenden (1999), "learner beliefs" is an interchangeable term for metacognitive knowledge. Regarding that, Flavell (1979) described, "metacognitive knowledge consists primarily of knowledge or beliefs about what factors or variables act and interact in what ways to affect the course and outcome of cognitive enterprises." Based on the definition, students' beliefs in a language learning context are what they believe affects their successful language learning. Therefore, students' beliefs or metacognitive knowledge about language learning can be connected to their learning strategies.

As for the connections between metacognitive knowledge and metacognitive strategies, Wenden (1999) stated that they are separate and distinct yet complementary components of the broader notion of metacognition. She defined metacognitive knowledge as information about language learning that learners acquired, and metacognitive strategies as general skills that learners manage, direct, regulate, and guide their learning. Thus, if, for example, a student believes that

repeating a word after a teacher is not an effective way of learning a language, they would not do so during a class, which might cause unsuccessful or tardy learning.

In order to empirically investigate the relationships between students' beliefs and their outcomes, Wang et al. (2009) studied relationships between 45 L2 Chinese learners' beliefs and their scores of course-end achievement tests, and found that students' beliefs about confidence and self-efficacy, and their course-end performance positively correlated (r = .42, p = .01). Moreover, Fahmimroah et al. (2017) studied relationships between students' beliefs towards plurilingualism and their academic achievement. With 180 samples of university students at English department in Indonesia, they found a positive significant correlation between the two variables (r = .563, $p < .05$). Their studies indicate that there seems to be relationship between the students' beliefs about language learning and their achievement.

As shown in the studies above, if students' beliefs or metacognitive knowledge about language learning, and their learning outcomes are somehow related, teachers need to know their beliefs that are substantially related to their learning materials. For example, in the case of conversation homework with a peer, it is worthful for teachers to learn students' beliefs about conversation partners' native speaker status. If students believe that conversating peers would not be very beneficial for their language learning, they might not earnestly engage in the homework. However, even if students have such beliefs, teachers can work on changing the beliefs if they find out about them. In this sense, it is vital to explore students' beliefs about conversation partners' native speaker status when outside-a-classroom conversation homework is given.

Research Questions

Inspired by the popularity of oral proficiency in language learning and teaching, the present research will address the following research questions (RQ1 & RQ2). The other research questions (RQ3, RQ4, RQ5 & RQ6) are related to the first two questions, and they examine educational appropriateness of the new intervention. RQ4, 5, and 6 in particular will explore L2 Japanese learners' beliefs about conversation partners.

RQ1. Does weekly video-synchronous conversation practice with a conversation partner help L2 Japanese learners improve their oral proficiency? If so, to what extent?

RQ2. If weekly video-synchronous conversation practice helps improve L2 Japanese learners' oral proficiency, is there any difference between having a peer conversation partner and having a native conversation partner in improving L2 Japanese learners' oral proficiency?

RQ3. What are the L2 Japanese learners' reactions toward weekly video-synchronous conversation practice?

RQ4. What are the L2 Japanese learners' beliefs toward having a peer conversation partner in weekly video-synchronous conversation practice?

RQ5. What are the L2 Japanese learners' beliefs toward having a native conversation partner in weekly video-synchronous conversation practice?

RQ6. What are the L2 Japanese learners' beliefs about conversation partner?

Summary

This chapter briefly discussed backgrounds and motivations of this study including high demands for improvement of oral proficiency in language learning, and connections between learners' beliefs and learning achievement. Research questions were presented at the end.

CHAPTER 2. LITERATURE REVIEW

In this chapter, first, some hypotheses about second language acquisition, and importance of communicative activities in second language acquisition will be discussed. Subsequently, two studies that examined the effectiveness of conversation practice via video-synchronous conferencing tools will be reviewed. Second, characteristics of conversation between non-native speakers, and its beneficial points for learners will be presented. Third, students' beliefs about language learning, especially about peer interaction will be discussed. Finally, discussions about an oral proficiency instrument that was used in the present experiment will be given.

Communicative Activities

Communicative activities for language learning

Today, oral proficiency has received more attention than before in foreign language education. As presented in Ch.1, ACTFL clearly gives weight to oral proficiency, requiring K-12 foreign language teachers to be orally proficient at certain levels at OPI so they can conduct classes in the target languages (ACTFL, n.d.). In the development of the field of SLA, a number of research studies from the past few decades have attempted to find theories that can explain the process of language acquisition.

For example, Krashen (1985) advocated the Input Hypothesis, which claims that a process of second language acquisition is through comprehensible input ("i +1") and production in the target language gradually emerges just as babies acquire their first languages. Further, Swain (1985) suggested the Output Hypothesis, which describes various types of output (e.g., a casual/formal conversation with a native speaker or a collaborative work with peers) are useful in second language acquisition because learners can learn from their language production by reflecting on their own output, reprocessing unaccepted output and modifying it. Long (2015) proposed the Interaction Hypothesis, which expounds that negotiation for meaning during communication breakdown draws learners' attention to linguistic code features, and because of that, explicit learning occurs.

Guided in part by these hypotheses, language teachers today make a great deal of effort in making the classroom activities more communicative with a focus on authentic language use rather than traditional grammatical instruction. Despite the teachers' best efforts to provide students with as many opportunities as possible to participate in communicative activities during class, class time and types of communicative activities that can be done in class are never enough, especially when the class is embedded in an institutional system. Indeed, even though foreign language teachers reported that they spend most time on speaking during class time among the four skills, they still strongly desire to make students practice speaking outside a classroom (Fukada & Wei, 2012). In terms of types of communicative activities in a classroom, students are only exposed to communications in the target language either with the instructor or peers on chosen topics, especially in the case of foreign language learning. Outside-the-classroom communicative activities such as casual conversation practice with native speakers or peers would offer a possible solution to these limitations.

There is another merit of outside-the-class communicative activities for students. According to Swain (1993), students should engage in extended discourse outside of language class where they have to push their linguistic knowledge to the fullest because students' language production elicited by a teacher during a class tends to be short and syntactically simple. So, in order to become a competent user of the target language who can use it beyond classroom discourse with their teacher and peers, it is necessary for students to practice using the target language outside the classroom, utilizing all the syntactic, lexical, and phonological knowledge that they learned in class.

Engaging in extended discourse outside of a classroom not only pushes students to use their linguistic knowledge to the fullest, but also gives them an opportunity to be exposed to a different type of language use from that in a classroom. Comparing conversations between a native speaker and a non-native speaker in a second language classroom with those observed outside a classroom, Long (1983) found that in an elementary-level language classroom, a native speaker's (a teacher's) speech included a lot of display questions and here-and-now topics that required fewer negotiations of meaning. On the other hand, conversations between a native speaker and a non-native speaker outside the classroom included a lot of referential questions that required a lot of negotiations of meaning (Long, 1983). Supported by these findings, in-class interaction does not seem sufficient for students to be able to communicate successfully in the real world using the target language.

His analysis was limited to classroom language use in an elementary-level language classroom, but language teachers tend to control their language more or less unless students are very advanced. From this point of view as well, students need to practice using the target language outside the classroom to be able to use the language in a communication.

Communicative activities via video-synchronous communication tools

In an effort to overcome these limitations and difficulties that foreign language learning faces, and to benefit from communicative activities outside of a classroom, some researchers have turned to the recent technological development which has enabled us to communicate with people at a distance (e.g., Skype, Zoom or Facebook) (Kato, et al., 2016; Saito & Akiyama, 2017; Spring et al., 2019; Turhune, 2016). This is definitely an exciting enhancement for foreign language learning and teaching. Today, some studies have investigated effective and practical ways to incorporate such recent technology into language pedagogy.

For instance, Kato, et al. (2016) reported on a semester-long online language exchange program via Skype. In their study, participants were 38 American learners of Japanese in the U.S. (26 in the experimental group and 12 in the control group), and 37 Japanese learners of English in Japan (26 in the experimental group and 11 in the control group). The participants in the experimental groups were engaged in regular Skype language exchange projects with a native speaker (treatment), while the control group did not. The participants had completed a short survey about their hobbies, available times for Skype meetings, and area of study at school before the program began. Then, the program leaders paired students in the American experimental group with the Japanese counterparts who had the most in common including meeting availability. The students were instructed to have two 30-minute conversation sessions each week for 15 weeks in which they had to speak in their native language for 15 minutes and in their target language for another 15 minutes per session. The participants were to discuss assigned topics in one session, and to engage in free conversation in another session each week. Their study employed a pre-post design, and evaluated the participants' improvement in oral proficiency using these two objective measures: speech rate and mean utterance length.

In terms of speech rate, both experimental groups made gains, while neither of the control groups showed significant improvement. Their study also found that there was a significant

relation with a large effect size between participation in the Skype partner program and an increase in speech rate for both experimental groups.

For mean utterance length, they reported that the American learners of Japanese, both experimental and control groups, made significant improvement although a larger effect size was found in the experimental group (d = 1.44) than in the control group (d = .5). They noted that the significant improvement of the American control group might have been a result of in-class instruction on speaking.

As for the Japanese learners of English, the experimental group was reported to have made significant gains with a large effect size in mean utterance length. Meanwhile, the control group showed a decrease in mean utterance length that was not statistically significant. They reported strong relations with large effect sizes between participation in the program and increase in mean utterance length for both experimental groups.

On the basis of these results, Kato et al. (2017) concluded that a video-synchronous computer-mediated language partner program pairing native speakers learning each other's languages is beneficial for learners to improve their oral proficiency, particularly when integrated into an ongoing course and when structured to support meaningful communication. However, it is unclear from their paper whether or not they succeeded in measuring exclusive effects of video-synchronous conversation practice; their control group of American learners of Japanese did not seem to take the same class as the experimental group did, which causes ambiguity of the improvement if it was brought about by the practice or the course instruction.

In addition, their study "examined" the effectiveness of a language exchange program using native speakers as interlocutors, but it did not address if the partner's native speaker status carried significance. It might be similarly effective even when peers are paired and used as interlocutors in online conversation practice. If so, paring peers would be more efficient than a language exchange with native speakers in the distance, because it would reduce the amount of teachers' work such as finding a corroborating institution. It can be more convenient for students as well in terms of scheduling; in the case of peer conversation partners, they should be in the same time zone.

Likewise, Saito and Akiyama (2017) examined whether or not virtual face-to-face interaction via video-conferencing tools with native speakers has the same effect as in-person face-to-face interaction with native speakers in terms of enhancement of learners' oral ability. Their

experiment adopted a pre-post design, and their participants were 30 Japanese undergraduate students (freshmen and sophomores) majoring in business at a university in Japan (15 in the experimental group and 15 in the control group). The participants in the experimental group participated in weekly virtual conversation practice with native speakers, while those in the control group were engaged in vocabulary/grammar activities. At the time of the project, all students were, as per a university requirement, enrolled in approximately three hours of EFL lessons per week taught by Japanese instructors, which mainly focused on listening and reading activities without many opportunities to produce language. Also, they noted that the participants had studied English in EFL classrooms for six years prior to the project, typically through grammar-translation methods, and had few opportunities to practice conversation in English.

Before they had started the program, Japanese students in the experimental group were paired with native speakers of English in the United States who were studying Japanese at a university. Their study differed from that of Kato et al (2016) in that they did not examine the improvement of the oral proficiency of the American students. The American students were trained to give explicit linguistic feedback (e.g., pronunciation, vocabulary, and grammar) as a form of recast during the sessions when the Japanese learners' utterance in English hindered successful comprehension. Given a task to work on together, each pair interacted for 60 minutes in weekly sessions for nine weeks: the first half of the session in English and the second half of the session in Japanese. In order to measure the effectiveness of their new intervention, they examined the students' improvement of English skills in various domains such as global impression (comprehensibility and accentedness), pronunciation (segmentals, word stress and intonation), fluency (speech rate), vocabulary (appropriateness and variation), and grammar (error ratio of verbs, nouns, and articles). They utilized seven picture description tasks as pre- and post- oral proficiency tests.

Having conducted nine sessions throughout a semester, their study found that the Japanese EFL learners had significantly improved their English skills in four domains out of 11 compared to the control group: comprehensibility, fluency, vocabulary variation, and error ratio of verbs and articles. From their results, Saito and Akiyama (2017) concluded that L2 learners can improve pronunciation, fluency, vocabulary, and grammar with primary attention devoted to maintaining successful communication with native speakers. The participants in their study were supposed to have practiced the target language much less in total (270 mins) than those in the experiment of

Kato et al (2016) (450 mins). Despite that fact, the participants significantly improved their oral skills in some domains. The feedback from the interlocutor might have somewhat influenced this result, but how much it affected the improvement was not reported in their study. In addition, similarly to the study of Kato et al (2016), their study did not address the question of whether or not it is crucial to have a native speaker as a conversation partner to make significant improvement. Further, the participants in their study were inexperienced Japanese learners of English, who might have had much more room for growth. That said, the same effects could not be expected on Japanese learners in the U.S. that are instructed in a different way (i.e. communicative approach) from English learners in Japan.

As these studies show, virtual conversation practice on video-synchronous communication tools seems to have similar effects as those of in-person, and outside-a-classroom conversation practice with native speakers using such tools help learners improve their oral proficiency. However, these studies cannot be legitimately generalized to different types of learners; the study of Kato et al. (2016) remains questionable whether they properly measured the effectiveness of the practice on American learners of Japanese, and that of Saito and Akiyama (2017) is limited to inexperienced learner with extremely few exposers to the target language aurally and orally. Moreover, they did not investigate the importance of interaction partners' native speaker status, and similar degrees of improvement might be observed in virtual conversation practice with peers.

The results of these studies are consistent with the claims of the aforementioned theories that learners need interaction that generates comprehensible input and output to develop their oral proficiency, and that learners need to practice using the target language not only inside a classroom but also outside to be fluent in the language. Video-synchronous communication tools expand these opportunities for learners and assist their language learning process. Therefore, the present study will empirically examine the following research questions to investigate the effectiveness of regular video-synchronous conversation practice with a peer or a native speaker, and students' reaction toward the practice.

RQ1: Does weekly video-synchronous conversation practice with a conversation partner help L2 Japanese learners improve their oral proficiency? If so, to what extent?

RQ2: If weekly video-synchronous conversation practice helps improve L2 Japanese learners' oral proficiency, is there any difference between having a peer conversation partner and having a native conversation partner in improving L2 Japanese learners' oral proficiency?

RQ3: What are the L2 Japanese learners' reactions toward weekly video-synchronous conversation practice?

Conversation Practice Between Non-native Speakers

In recent years, pair- and group- work are widely used activities in language classes. These interactions are between non-native speakers, which might have some characteristics that are different from those between native speakers. In order to capture features of conversations between non-native speakers, between native speakers, and between a non-native and a native speaker, Varonis and Gass (1985) analyzed conversations between speakers with various linguistic backgrounds. They observed informal conversations between different types of speakers: fourteen conversational dyads between non-native speakers, four conversational dyads between native and non-native speakers, and four conversational dyads between native speakers. All the dyads were composed of speakers who had not met before the research, and all the non-native speakers were native speakers of either Spanish or Japanese who were students at the English Language Institute of The University of Michigan at the time of the research. Analyzing the discourses of each dyad by categories that they created for different types of conversation flows (e.g., non-understanding, misunderstanding, ignore, comment upon it, trigger as question, and so on), they found that the number of occurrences of "non-understanding" in the conversations between non-native speakers was statistically greater than that of the other dyads. The conversations between non-native speakers did not flow as smoothly as the other dyads with native speakers because of their language low proficiency, which resulted in more time spent on negotiation of meaning (Varonis & Gass, 1985). They concluded that the more non-native speakers are involved in a dyad, the more time interlocutors will spend in negotiation of meaning.

Varonis and Gass (1985) also discussed why non-native speaker dyads spent more time for meaning negotiation than did the other dyads. They concluded that non-native speakers feel less

threatened to show their incompetency to other non-native speakers than to native speakers, which makes them actively involved in negotiation of meaning.

Similarly, having reviewed literature on group work and interlanguage talk, Long and Porter (1985) reported that the amount of speech and the range of language functions (rhetorical, pedagogic, and interpersonal) practiced by students are likely to increase in group work, compared to that in teacher-led instruction. Also, in terms of accuracy of speech, students seem to maintain the same level of accuracy in supervised and unsupervised environments, so it would not be true to say that lower quality is the price to be paid for a higher quantity of practice.

From these findings, conversations between non-native speakers seem to have some merits for learners. First, peer conversations are likely to generate more negotiations of meaning than those including a native speaker. That is because they often go into communication failure resulting from their linguistic weakness, and when that happens, they feel comfortable working on resolving the problems with other non-native speakers. In line with the Interaction Hypothesis (Long, 2015), conversations between non-native speakers that markedly contribute to negotiations of meaning are more favorable than those with a native speaker. Second, students can practice various language functions in peer interaction, which was reported to differ from those between students and a teacher. What is more, the same level of students' speech accuracy as students-teacher interaction is obtained in peer interaction.

Supported by these findings, learners would benefit more from conversation practice with other non-native speakers than that with native speakers. It will be a practical and realistic way for students to practice speaking in a more comfortable environment outside the classroom and will be a feasible tool for teachers as a form of speaking homework. Though the previous research shed light on the merits of interactions between non-native speakers, it is not known yet whether or not there would be a difference in effectiveness between conversation practice with a native speaker and a peer. Thus, RQ2 shown in the previous section will examine the issue as well.

Students' Beliefs About Peer Interaction

Peer interactions in a form of pair work or group work are used by teachers to increase the opportunity for students to practice speaking a target language, but it should be interesting and important to hear students' opinions on whether or not they think peer interactions are a useful and worthful way to improve their language skills, especially the speaking skill.

It is possible that students believe that peer interaction is a less effective way to improve the target language than interaction with a native speaker of the target language because peers are merely non-native speakers just like them. It is not unusual to hear a complaint from students that they prefer a native speaker as their foreign language teacher rather than a non-native speaker, however proficient they might be. In fact, a professor who is in charge of English classes for international graduate teaching assistants in a U.S. university shared that a student in one of these English classes came to her and complained why they had a non-native English teacher, instead of a native one (A. Ginther, personal communication, 2020).

Students' beliefs in language learning

In order to study beliefs about language learning, Horwitz (1985) developed the Beliefs About Language Learning Inventory (BALLI), and it is used to assess both teachers' and students' beliefs about language learning. There are a teachers' version that contains 27 items, and an ESL students' version that contains 34 items. Both versions employ a Likert scale, and are designed to elicit subjects' beliefs in four and five major areas, teachers' version and students' version, respectively: foreign language aptitude, the difficulty of language learning, the nature of language learning, learning and communication strategies, and motivations (only for the students' version).

Horwitz (1987) pointed out that it is important to study students' beliefs about language learning because beliefs can affect students' language learning strategies, which in turn affect their learning outcomes. She explained in an ESL context, "a student who believes, for example, that one must never say anything in English until it can be said correctly will probably avoid speaking most of the time" (p.120). The belief of this student would presumably have a negative influence on their language learning, and it would be helpful for their instructor to know about such beliefs; the instructor would be able to approach the student in a special way to encourage them to speak even before they can say something perfectly in English.

L2 Japanese learners' beliefs about peer interaction

Motivated by Horwitz's inventory, research about a wide range of teachers' and learners' beliefs has been actively carried out in the field of Japanese pedagogy as well. Among the research, some studies about L2 Japanese learners' beliefs about peer interaction will be reviewed below.

Based on the students' version of BALLI, Abe (2009) created an original survey to examine beliefs of L2 Japanese university students in Spain. He surveyed 98 university students in Madrid who were learners of Japanese at that time. The survey deployed a five-point Likert-scale (1=strongly agree to 5=strongly disagree), and revealed that most of the students believed that peer interaction is meaningful; the average rating for the question, "speaking Japanese with other learners in a classroom does not help my learning." was 4.1.

Similarly, Matsumoto (2020) researched 274 Costa Rican L2 Japanese learners' beliefs about language learning who were studying Japanese either in a university or in a language school at the time of the research. Like Abe's survey, Matsumoto's survey also deployed a five-point Likert-scale (1=strongly agree to 5=strongly disagree), and contained the same question about peer interaction as Abe's ("speaking Japanese with other learners in a classroom does not help my learning"). The result is analogous to Abe's, and the majority of the students disagreed that peer interaction in a classroom was worthless; the mean rating was 4.26.

Furthermore, Kishi (2020) researched 66 L2 Japanese learners in Egypt who majored in Japanese in a university. His survey included the same question as Abe's and Matsumoto's ("speaking Japanese with other learners in a classroom does not help my learning"), and reported that the majority of the students agreed that peer interaction is worthful; 78.8 % disagreed with the question (a four-point Linkert scale without "neither agree nor disagree" deployed).

As shown above, the students in the studies from three regions seem to have a positive attitude toward peer interaction in a classroom.

Students' beliefs about native/non-native status

The three studies above elicited students' beliefs about peer interaction in a classroom. These students may have responded this way thinking that peer interaction in a classroom is helpful for their learning if the alternative is no interaction with anyone in class; in a typical language class, peer interaction is often the only way to practice speaking with an interlocutor. In other words, the students' beliefs elicited by the aforementioned studies were about peer interaction, not about native/non-native status. Therefore, if learners are given a choice between a peer or a native speaker of the target language, as their interaction partner, their answer might be different; they might believe that interaction with a native speaker of the target language is more beneficial for improving their speaking skill than with a peer, and vice versa. By making them choose between

a peer or a native speaker, learners' beliefs about conversation practice partners in language learning will be more clearly revealed. Also, positive experience of peer interaction might change their beliefs. Therefore, this study will investigate the following research questions.

RQ4: What are the L2 Japanese learners' beliefs toward having a peer conversation partner in weekly video-synchronous conversation practice?

RQ5: What are the L2 Japanese learners' beliefs toward having a native conversation partner in weekly video-synchronous conversation practice?

RQ6: What are the L2 Japanese learners' beliefs about conversation partners?

Oral Proficiency Test Instrument: Elicited Imitation

Since this study quantitatively examines students' improvement of oral proficiency affected by regular online conversation practice, a quantitative oral proficiency measurement is used. Elicited Imitation (EI) is a practical method to measure the oral proficiency of L2 learners. In the EI task, the subject is requested to repeat the stimulus sentences as exactly as possible after a few seconds pause. The stimuli are aural ones and recorded by a native speaker. Then, the accuracy of the repetition is examined to measure the subject's oral proficiency.

What EI measures

EI is a language test to measure the subjects' wide range of linguistic knowledge and acquisition of aural/oral skills of the target language. EI has been used widely in the field of first and second language acquisition in various ways. For example, EI can be used to measure learners' specific grammatical knowledge, and learners' global oral proficiency. Yan et al. (2016) concluded from their meta-analysis that EI is a sensitive language assessment of global language proficiency and specific structures, and it can be used to measure the effectiveness of instructional interventions. They also found that EI measures global constructs better than it does specific grammatical knowledge or phonological acquisition.

Features of EI

EI has some important features to make it a valid language test. One of them is a few seconds pause before the subject is to start repeating. Because of the simple procedure used in EI, which is to repeat what you just heard, there may be a concern that the subject only imitates the sounds from their short-term memory without understanding the content of the sentence, and EI does not measure the subject's oral proficiency but their memory capacity (Vinther, 2002). In order to ensure that it measure the subject's oral proficiency, not their memory capacity, a prompt for repetition is delayed long enough to prevent the subject from accessing their short-term memory. Thus, in order to successfully repeat a stimulus sentence, the subject must decode it first and then reconstruct it using their linguistic knowledge (Vinther, 2002). This way, EI measures the subject's aural/oral skills, accuracy of comprehension and production, and proficiency.

To ascertain the importance of comprehending stimulus sentences for successful repetition, Vinther (2002) conducted an experiment in which the subjects were instructed to repeat meaningful and nonsense sentences of varying lengths. In the experiment, six native Spanish speakers were requested to repeat four meaningful sentences and six nonsense sentences which differed in length: three meaningful sentences with 26 syllables and one with 16 syllables, and three nonsense sentences with 26 syllables and three with 16 syllables. Just like the meaningful sentences with correct Spanish spellings, the nonsense sentences were made up of Spanish-sounding nonsense words, and they are read out with natural Spanish intonation. The results showed that the native Spanish speakers accurately repeated all the meaningful sentences and the nonsense sentences with 16 syllables, but not the nonsense sentences with 26 syllables. As this experiment revealed, if the stimuli are short enough to be stored in the subject's short-term memory, it is possible to repeat them accurately without understanding the meaning. However, if the stimuli are beyond their memory capacity, the subject must decode them first and reproduce them. Thus, stimulus sentences for EI tasks need to be long enough to challenge the subject's memory capacity.

Furthermore, Yan et al. (2016) found that stimulus sentences that vary in length rather than fixed length throughout the task enhance the sensitivity of EI. Therefore, in addition to the requirement that the stimulus sentences need to exceed the subjects' capacity of short-term memory, they need to vary in length.

Scoring

There are many ways to score an EI item: e.g., either count correct syllables or morphemes, whether the whole sentence is correct or incorrect (all or nothing), and partial scoring (5=perfect repetition, and 1=no repetition). There is also a scoring method that identifies a specific grammatical structure: if the target structure is successfully repeated, it is marked as + ignoring other erroneous repetitions (Schwartz & Daly, 1976). In order to determine the best scoring method that can measure the subjects' acquisition level, Miller and Chapman (1975) investigated which of the following three factors correlates the most strongly with the difficulty of stimulus sentences: the number of morphemes, the number of words, and lexical density. Scores should best reflect subjects' acquisition level, so it should co-vary with the sentence difficulty; the more difficult the sentence is, the lower the score should be. Their experiment demonstrated that the number of morphemes rather than the number of words had a stronger correlation with sentence difficulty, and they concluded and suggested that a morpheme might be the most appropriate unit in scoring EI.

In addition, Yan et al. (2016) concluded that a more refined scoring method than binary makes EI a more sensitive language test to measure subjects' proficiency levels. Thus, again, a morpheme is the most appropriate scoring unit in EI.

EI's validity

As for the external validity of EI as an oral proficiency test, Bowden's study (2016) explored correlation between EI and an external oral proficiency test. In the study, using L2 Spanish learners with various learning experiences and backgrounds, she compared the learners' EI scores with their scores on the Simulated Oral Proficiency Test (SOPI). According to Stansfield and Kenyon (1992), SOPI is a semi-direct oral proficiency test that is based on the ACTFL guidelines, and it elicits the examinees' spoken language using steps similar to the Oral Proficiency Interview (OPI); starting with simple personal background questions posed on a tape in a simulated initial encounter with a native speaker of the target language, test takers demonstrate their speaking ability on various tasks associated with different levels of the scale. After the test is completed, the recorded performance is scored by trained raters following the ACTFL proficiency scale. Although SOPI differs from OPI in terms of the test protocol, Stansfield and Kenyon (1992) concluded that the two tests are close enough in the way they measure general speaking proficiency that they may

be viewed as parallel tests delivered in two different formats. Using the SOPI scores as an external oral proficiency test to be compared with the EI scores, Bowden's study found a very high correlation ($r = .911$, $p < .01$) between the two tests, and she concluded that EI is a valid and useful tool that is comparable to SOPI and can discriminate among a broad range of L2 learners when designed and administrated appropriately.

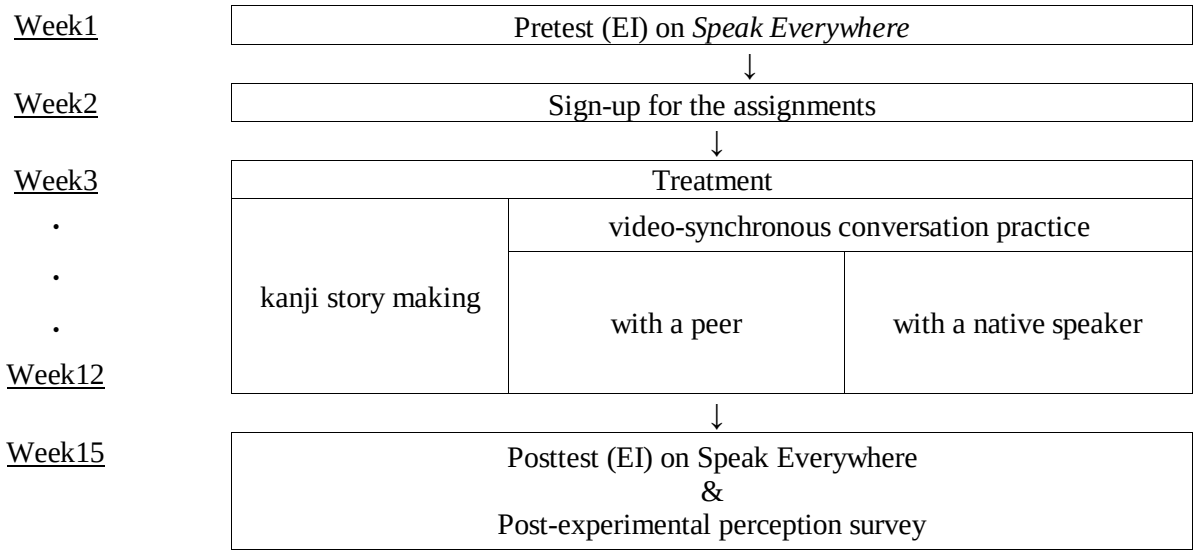
Similarly, TrueNorth, a language testing company, examined the relationship between test takers' scores on their Spanish EI test and those on ACTFL's Oral Proficiency Interview Computer (OPIc), and found a high positive correlation between the two sets of scores ($r = .903$, $p < .001$) (TrueNorth, 2018).

Summary

This chapter briefly presented the importance of communicative activities for language learning, and previous research about video-synchronous conversation practice. Moreover, features of conversation between non-native speakers, and the benefits of these types of conversations for learners were reviewed. Further, students' beliefs about language learning with a special focus on peer interaction were discussed. Finally, the oral proficiency test instrument employed in this study was presented.

CHAPTER 3. METHODS

This study examined the effectiveness of regular video synchronous conversation practice in helping L2 Japanese learners improve their oral fluency. Participants were 31 third-year Japanese learners at a public university in the Midwestern United States, and they participated in a semester-long pre-post design experiment in the Fall 2020 semester. They were divided into three groups: a control group and two experimental groups. The control group participated in a non-speaking assignment (kanji story making), and the two experimental groups participated in weekly video-synchronous conversation practice with a peer or a native speaker of Japanese. They chose an assignment from the two at the beginning of the semester. An Elicited Imitation (EI) test instrument was used to measure the learners' pre- and post- oral performance. The schedule of the experiment is shown below. Students were to choose the assignment type by signing up in an online sheet.



Japanese Course (JPNS 30100)

The course chosen for the experiment was a 3rd-year 1st semester Japanese course at a public university in Indiana, the United States, which is the fifth level course at the present institution. The course had two sections that were taught by the same instructor and covered the

first five chapters of the course textbook *QUARTET: Intermediate Japanese Across the Four Language Skills I* (Sakamoto et al., 2019). The course met three times a week for 50 minutes per class in person or synchronously online. Around seven class sessions were devoted to covering each chapter that consisted of two kinds of instruction; one was teacher-led lectures that aimed to teach new grammar and to deepen students' understanding about textbook readings, and the other was students-led classes in which students discussed the readings and engaged in conversation practice in small groups or in pairs. As an extra credit activity, "Japanese Corner" was offered every other week where students enjoyed free conversation in Japanese with the other participants and the instructor for an hour. In addition, there were other opportunities for students to practice speaking: e.g., two individual interview tests with the instructor, three chapter-end video-making projects, an animation voice-over group project, an individual oral presentation, reading aloud homework, and so forth. Overall, the course was designed with an extraordinary focus on oral skills:

Participants

L2 Japanese learners

Thirty-one learners of Japanese who were enrolled in the course mentioned above participated in a semester-long experiment as a part of their coursework.

The participants were all undergraduate students (21 male and 10 female), and their majors widely varied. Three of them had Japanese as one of their double- or triple-majors. Four students were freshmen (12.9%), four were sophomores (12.9%), 12 were juniors (38.7%), and 11 were seniors (35.5%). Most of them were L1 English speakers, and the others were seven L1 Chinese speakers, two L1 Korean speakers, one L1 Thai speaker, and one L1 Portuguese speaker. Four students were heritage speakers of Japanese, and five had a study abroad experience in Japan for a few weeks up to one year. Nineteen of them had taken lower courses at the institution as prerequisites, and 12 of them learned Japanese in high school and were placed into the 3rd-year level by a placement test.

This level was chosen for the study for two reasons. First, previous research (Spring et al., 2019) recommends that students have had at least five semesters of language instruction before utilizing a native-speaker as a conversation partner in video-synchronous mediated communication.

Second, this course had a sufficient number of students to form three groups needed for the present research. The learners had had four semesters of university-level language instruction or equivalent knowledge before participating in the research.

Native speakers of Japanese

Two groups of native speakers of Japanese participated as conversation partners: L1 Japanese university students and L1 Japanese volunteers. Eighteen L1 Japanese students who were enrolled in a Japanese private university and taking a Japanese language pedagogy course at the time of this research participated as a part of their coursework for four weeks. They were sophomores, juniors, and seniors at the time of this research (two male and 16 female). For the remaining six weeks, native speakers of Japanese were recruited as volunteers via the present author's personal connection. In total, 11 L1 Japanese volunteers participated that were aged between 17 and 26 (six male and five female).

Materials

Remembering the Kanji (Heisig, 2011)

For the learners that chose the assignment of kanji story making, this book was provided as a reference, from which they could select kanji to work on. The kanji story was a kanji-learning method proposed by Heisig (2011) for learners of Japanese to easily memorize and recall kanji. The method introduces a way to associate meanings of various kanji with the forms. For this assignment, it was their choice whether or not to use the book as a reference and translate the story written in English into Japanese adding some original ideas; students were also welcome to create their own kanji story from scratch.

A video-synchronous communication tool

Video-synchronous communication tools for the conversation practice needed to meet the following two requirements: functions to maintain at least 30-minute sessions continuously and to record the video and the audio of an entire session. The learners were allowed to use their preferred communication tools as long as they met these two requirements (e.g., Skype or Zoom)

Conversation topics and conversation-generating questions

For each conversation practice session, a topic and 8 to 10 conversation-generating questions were provided to encourage active interactions between the participants. Topics were selected based on the textbook contents, and conversation-generating questions were created based on conversation topics. The learners were given all conversation topics and conversation-generating questions at the beginning of the semester, and native speakers of Japanese were provided with them before they began the conversation sessions with the participants.

Post-experimental perception survey

A survey was administrated at the end of the semester in order to collect the participants' demographic information, and to elicit their overall reactions to the assignments (kanji story making and conversation practice), and beliefs about conversation partners. There were three versions which employed multiple-choice, a five-point Likert scale and text entries. All three surveys contained nine demographic-related items, and 13 reaction-related items. As for the belief-related items, there were two items for the conversation groups (Peer Group and Native Group), and one for the Kanji Group. The original surveys are listed in the appendix. The survey was created and administered anonymously through an online platform called *Qualtrics*. The learners were to complete the survey right after the posttest.

Testing Material

All the participants took a pre- and post-test in the form of Elicited Imitation (EI) test instrument as oral proficiency tests. The EI task included 20 Japanese sentences controlled for grammatical structure, vocabulary, and length. All the grammar and vocabulary were limited to those of the second-year level at the institution, and they should have been familiar to the participants. The 20 sentences varied in length; the shortest one consisted of 11 morphemes (e.g., すぐ戻れるなら窓は開けたままでもいいよ), and the longest one 19 morphemes (何時間もかけて終えた宿題を寮の部屋に忘れてきてしまった). All the sentences can be found in the appendix.

In order to divide the sentences into morphemes, which is challenging even for native speakers because of the high complexity of Japanese grammatical structure, the morphological analyzer application mecab v.0.996 together with the Japanese lexicon unidic v.2.1.2 was used. The sentences were read aloud and recorded by the present researcher, a 26 years old female L1 Japanese speaker, and it was confirmed that a highly proficient non-native speaker of Japanese was able to repeat them without any problems. The pre- and post-tests were administrated through *Speak Everywhere* (Fukada 2009; 2013); the learners were instructed to complete the tests while being proctored on Zoom by the present researcher.

Procedure

Pretest and group assignment

In Week 1, the participants took the pretest. In both pre- and post-tests, a morpheme was used as a scoring unit. The present researcher rated each morpheme and counted the number of morphemes that were produced correctly. Synonyms and similar expressions were graded as incorrect in order to ensure objectivity throughout the items and subjects.

By the end of Week 2, learners chose an assignment from two different types: kanji story making or video-synchronous conversation practice. Both options were a part of their weekly homework and adjusted to be the same workload that can be completed within 30 minutes every week. Once they chose the assignment type, they had to stay with it throughout the semester. They worked on these assignments from Week 3 to Week12.

Treatment

Kanji-story making group (Kanji Group)

Thirteen learners who chose kanji story making were instructed to select an unlearned kanji every week and to make a short story about the kanji in Japanese incorporating its meaning and shape. Their kanji stories could include not only text but also images. They were to submit their kanji stories on the online learning platform of the university called Brightspace. Their kanji stories were shared with the whole class so the other learners including those who did not choose the kanji option could benefit from them.

Video-synchronous conversation practice groups (Peer Group and Native Group)

The remaining 18 learners who chose video-synchronous conversation practice were divided into two experimental groups according to their pretest scores. There were two experimental groups; one was a group that had a peer as a conversation partner (Peer Group), and the other had a native speaker of Japanese as a conversation partner (Native Group).The learners' pretest scores were sorted in descending order and the participants were divided into two experimental groups alternately; the highest student was placed into the Peer Group, the second-highest student was placed into the Native Group, the third-highest student was placed into the Peer Group, and so forth. The Peer Group consisted of ten students and they were paired with other classmates or coursemates. The Native Group consisted of eight students and they were matched with a native speaker of Japanese.

The Peer Group students were matched with a peer that had the closest pretest score to them from the top, and the partner stayed the same throughout the semester. The Native Group students were matched with a native speaker randomly by the researcher. They had a different conversation partner every week because of scheduling difficulty with the native speakers of Japanese. The Native Group students were informed of a conversation partner's name and email address by the end of Sunday every week and had to email them and schedule a meeting every week.

Both groups were instructed to hold at least a 30-minute conversation session via a video-synchronous communication tool every week. Conversation topics and conversation-generating questions were provided as supporting materials to help them maintain a conversation during the session. However, the participants were not limited to only talking about the given topics or questions and were given the freedom to develop their conversation. The learners were also allowed to use other functions equipped on the chosen communication tool such as screensharing or text-chat occasionally. The learners had to video-record the session and submit it on Brightspace.

Posttest and post-experimental perception survey

In Week 14, the participants took the posttest. The posttest was exactly the same as the pretest both in content and procedure. After having completed the test, each student was given a link to the survey and instructed to finish it.

Data Analysis

This study employed the Pre-Post Design, which is to examine the same participants at two or more levels repeatedly for one factor. The independent variable was groups (Kanji Group, Peer Group, and Native Group), and the dependent variables were students' pre- and post-test scores. A paired t-test was run to examine their improvement in oral proficiency, comparing the mean scores of the pre- and post-tests. A one-way ANOVA was used to examine if there were differences in each group's improvement. The statistical analysis was run via *IBM SPSS Statistics Version 26*.

CHAPTER 4. RESULTS

This chapter first presents the results of the data analysis of participants' pre- and post- oral proficiency tests (EI) to investigate whether or not weekly video-synchronous conversation practice improved their oral proficiency. If so, a further analysis will be presented to investigate whether or not the improvement of each group statistically differs from one another. Then, the results of the post-experimental perception survey will be presented.

Improvement of Oral Proficiency

The data analyzed for this research contained 30 participants out of the 31 total participants. One participant from the Peer Group (Conv.) was excluded from the data analysis because the participant's pretest was not properly proctored. First, descriptive statistics of pre- and post-tests of all three groups will be presented to show the characteristics of each group. Second, a comparison between pre- and post-test scores of each group will be presented to see if the students made gains in oral proficiency. Then, a comparison of gains among all three groups will be presented.

Descriptive statistics of pre- and post-tests

Descriptive statistics of participants' pre- and post-tests scores for each group are shown below. Tables 1-3 show the descriptive statistics for the Kanji Group, the Peer Group (Conv.), and the Native Group (Conv.), respectively. The score range is 0 to 283. No outliers were found in the pretest dataset of any of the groups (Figure 1).

Table 1. Descriptive Statistics of Kanji Group

		ScoresPre	ScoresPost
N	Valid	13	13
	Missing	0	0
Mean		104.69	133.77
Median		101.00	115.00
Mode		61	96
Std. Deviation		43.995	48.030
Skewness		0.685	0.356
Std. Error of Skewness		0.616	0.616
Kurtosis		-0.616	-1.622
Std. Error of Kurtosis		1.191	1.191
Minimum		57	76
Maximum		191	210

Table 2. Descriptive Statistics of Peer Group (Conv.)

		ScoresPre	ScoresPost
N	Valid	9	9
	Missing	0	0
Mean		134.33	165.56
Median		112.00	156.00
Mode		45[a]	52[a]
Std. Deviation		72.550	72.528
Skewness		0.856	0.043
Std. Error of Skewness		0.717	0.717
Kurtosis		0.099	-0.767
Std. Error of Kurtosis		1.400	1.400
Minimum		45	52
Maximum		273	269

a. Multiple modes exist. The smallest value is shown

Table 3. Descriptive Statistics of Native Group (Conv.)

		ScoresPre	ScoresPost
N	Valid	8	8
	Missing	0	0
Mean		148.50	175.63
Median		122.00	177.50
Mode		83[a]	91[a]
Std. Deviation		66.845	59.926
Skewness		0.827	0.105
Std. Error of Skewness		0.752	0.752
Kurtosis		-0.650	-1.065
Std. Error of Kurtosis		1.481	1.481
Minimum		83	91
Maximum		266	265

a. Multiple modes exist. The smallest value is shown

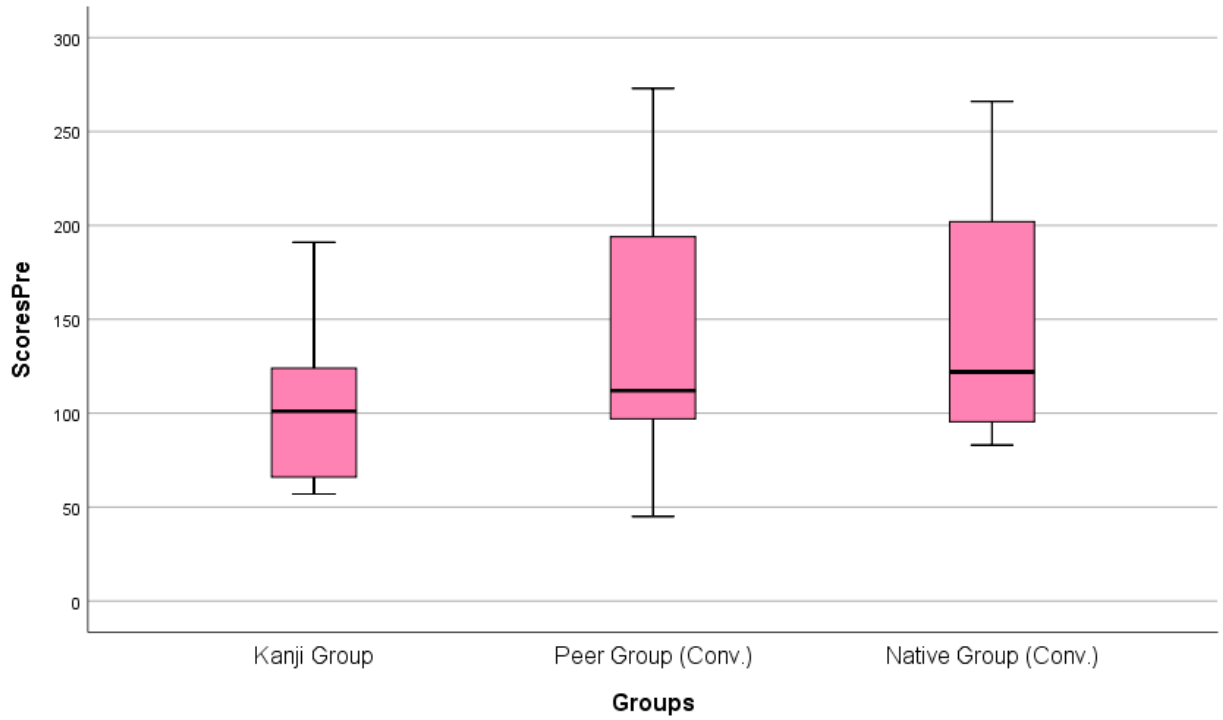


Figure 1. Pretest Scores by Groups

As shown in the tables and the figure above, the participants' pretest scores within each group varied to a large extent, especially within the two conversation groups.

Checking homogeneity of the three groups

Before starting the experiment, the three groups were examined by a one-way ANOVA to see if they were homogeneous. As prerequisites for a one-way ANOVA, Test of Homogeneity of Variances and Tests of Normality were run on the dataset. Tables 4 and 5 indicate that the dataset met the prerequisites to run a one-way ANOVA.

Table 4. Test of Homogeneity of Variances (pre-experimental checking)

		Levene Statistic	df1	df2	Sig.
Pretest Scores	Based on Mean	1.837	2	27	0.179
	Based on Median	0.602	2	27	0.555
	Based on Median and with adjusted df	0.602	2	19.928	0.557
	Based on trimmed mean	1.596	2	27	0.221

Table 5. Tests of Normality (pre-experimental checking)

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Groups	Statistic	df	Sig.	Statistic	df	Sig.
Pretest Scores	Kanji Group	0.169	13	.200*	0.905	13	0.156
	Peer Group (Conv.)	0.245	9	0.127	0.918	9	0.380
	Native Group (Conv.)	0.228	8	.200*	0.885	8	0.212

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Then, a one-way ANOVA was run to compare the pretest means of the three groups (Table 6). Although large within-group variations were observed, the three groups were not significantly different (p = .25) as shown in Table 6. So, the experiment proceeded with this grouping.

Table 6. ANOVA (pre-experimental checking)

Pretest Scores

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10561.097	2	5280.549	1.476	0.246
Within Groups	96612.769	27	3578.251		
Total	107173.867	29			

Pretest scores vs. posttest scores

After the posttest data had been collected, the scores of pre- and post-tests were compared one group at a time. A paired t-test was run to compare the means of the pretest and posttest. As shown in Table 7, all the groups scored higher on the posttest than the pretest, and the differences were statistically significant ($p < .00$, $p = .01$, and $p = .01$, Kanji Group, Peer Group (Conv.), and Native Group (Conv.), respectively)

Table 7. Paired Samples Test (pretest vs. posttest)

Groups			Paired Differences					t	df	Sig. (2-tailed)
			Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
						Lower	Upper			
Kanji Group	Pair 1	Scores Post - Scores Pre	29.077	25.395	7.043	13.731	44.423	4.128	12	0.001
Peer Group (Conv.)	Pair 1	Scores Post - Scores Pre	31.222	24.606	8.202	12.309	50.136	3.807	8	0.005
Native Group (Conv.)	Pair 1	Scores Post - Scores Pre	27.125	22.119	7.820	8.633	45.617	3.468	7	0.010

Table 8 shows the correlation between the participants' scores of the pre- and post-tests by groups. The correlations of all three groups were very high, which indicates that no ceiling effects occurred.

Table 8. Paired Samples Correlations

Groups			N	Correlation	Sig.
Kanji Group	Pair 1	ScoresPost & ScoresPre	13	0.851	0.000
Peer Group (Conv.)	Pair 1	ScoresPost & ScoresPre	9	0.942	0.000
Native Group (Conv.)	Pair 1	ScoresPost & ScoresPre	8	0.945	0.000

Figure 2 depicts the improvement of oral proficiency of all groups.

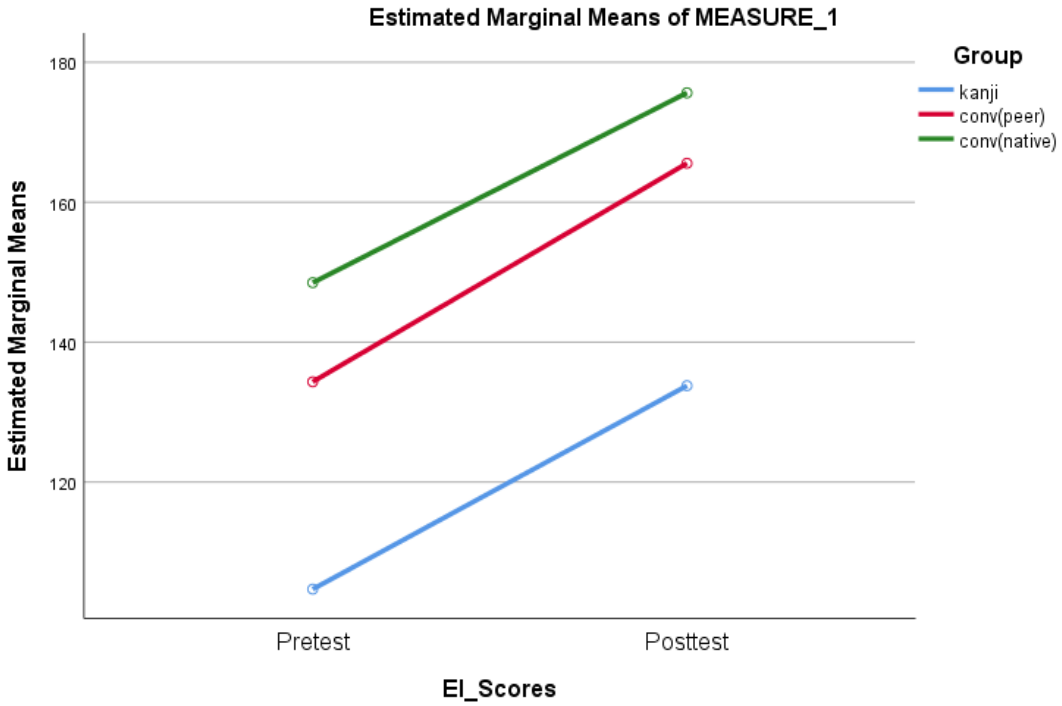


Figure 2. Gains by Groups

A comparison of gains among the three groups

In order to statistically compare the gains among the three groups, a further analysis was conducted. First, Table 9 presents the descriptive statistics of gain scores (posttest – pretest) of each group.

Table 9. Descriptive Statistics (posttest – pretest)

	Groups			Statistic	Std. Error
Score Difference	Kanji Group	Mean		29.0769	7.04333
		95% Confidence Interval for Mean	Lower Bound	13.7308	
			Upper Bound	44.4230	
		5% Trimmed Mean		27.1966	
		Median		26.0000	
		Variance		644.910	
		Std. Deviation		25.39508	
		Minimum		-5.00	
		Maximum		97.00	
		Range		102.00	
		Interquartile Range		19.00	
		Skewness		1.652	0.616
		Kurtosis		3.922	1.191
	Peer Group (Conv.)	Mean		31.2222	8.20193
		95% Confidence Interval for Mean	Lower Bound	12.3085	
			Upper Bound	50.1359	
		5% Trimmed Mean		31.3025	
		Median		26.0000	
		Variance		605.444	
		Std. Deviation		24.60578	
		Minimum		-4.00	
		Maximum		65.00	
		Range		69.00	
		Interquartile Range		46.00	
		Skewness		0.250	0.717
		Kurtosis		-1.287	1.400

Table 9 continued

Native Group (Conv.)	Mean		27.1250	7.82039
	95% Confidence Interval for Mean	Lower Bound	8.6327	
		Upper Bound	45.6173	
	5% Trimmed Mean		26.8611	
	Median		25.0000	
	Variance		489.268	
	Std. Deviation		22.11940	
	Minimum		-1.00	
	Maximum		60.00	
	Range		61.00	
	Interquartile Range		42.50	
	Skewness		0.239	0.752
	Kurtosis		-1.327	1.481

Similarly, Figure 3 depicts the gains of each group. One outlier was detected in the Kanji Group with a gain score of 97; the mean gain of the group was 29. Having consulted with the course instructor, the outlier student was not found to be a special case, so they remained in the dataset.

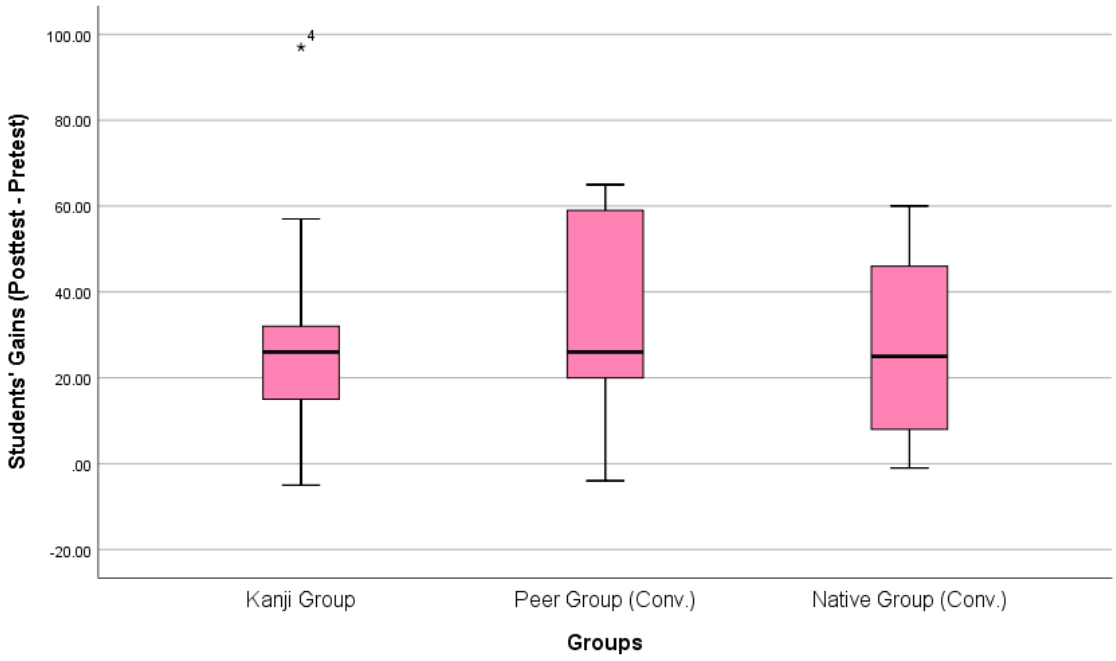


Figure 3. Gains of Each Group

Although Figure 3 illustrates seemingly similar gains among the three groups, a one-way ANOVA was run in order to examine them statistically. Before running a one-way ANOVA, Test of Homogeneity of Variances and Tests of Normality were run on the dataset as prerequisites to an ANOVA. Table 10 shows that the homogeneity assumption was met, but Table 11 shows that the normality requirement was not met.

Table 10. Test of Homogeneity of Variances (with the outlier)

		Levene Statistic	df1	df2	Sig.
Score Difference	Based on Mean	0.142	2	27	0.868
	Based on Median	0.079	2	27	0.924
	Based on Median and with adjusted df	0.079	2	23.652	0.924
	Based on trimmed mean	0.176	2	27	0.840

Table 11. Tests of Normality (with the outlier)

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Group	Statistic	df	Sig.	Statistic	df	Sig.
ScoreDifference	Kanji Group	0.254	13	0.022	0.855	13	0.033
	Peer Group (Conv.)	0.235	9	0.165	0.902	9	0.265
	Native Group (Conv.)	0.128	8	.200*	0.955	8	0.762

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

When the outlier was excluded, the two assumptions were met as shown in Table 12 and Table 13 below.

Table 12. Test of Homogeneity of Variances (without the outlier)

		Levene Statistic	df1	df2	Sig.
Score Difference	Based on Mean	1.730	2	26	0.197
	Based on Median	1.082	2	26	0.354
	Based on Median and with adjusted df	1.082	2	22.793	0.356
	Based on trimmed mean	1.729	2	26	0.197

Table 13. Tests of Normality (without the outlier)

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Group	Statistic	df	Sig.	Statistic	df	Sig.
ScoreDifference	Kanji Group	0.148	12	.200*	0.966	12	0.865
	Peer Group (Conv.)	0.235	9	0.165	0.902	9	0.265
	Native Group (Conv.)	0.128	8	.200*	0.955	8	0.762

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Since one of the assumptions was violated with the original data, two separate results of a one-way ANOVA (with and without the outlier) will be shown below. Table 14 and Table 15 present the results with the outlier (p = .94) and without the outlier (p = .7), respectively. Since the results are the same (i.e. both non-significant), the one with the outlier (p = .94) is treated as the final result for the analysis.

Table 14. ANOVA (with the outlier)

ScoreDifference

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	71.446	2	35.723	0.060	0.942
Within Groups	16007.354	27	592.865		
Total	16078.800	29			

Table 15. ANOVA (without the outlier)

ScoreDifference

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	314.101	2	157.051	0.371	0.694
Within Groups	11009.347	26	423.436		
Total	11323.448	28			

Lastly, Table 16 presents details of the post-hoc pairwise comparisons. The post-hoc pairwise comparisons are shown to emphasize that large standard error values are seen in each group, demonstrating once again that gains of individuals varied a great deal within the groups.

Table 16. Post-hoc pairwise comparisons (with the outlier)

Dependent Variable: Score Difference

Bonferroni

(I) Groups		Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Kanji Group	Peer Group (Conv.)	-2.14530	10.55836	1.000	-29.0951	24.8045
	Native Group (Conv.)	1.95192	10.94135	1.000	-25.9754	29.8793
Peer Group (Conv.)	Kanji Group	2.14530	10.55836	1.000	-24.8045	29.0951
	Native Group (Conv.)	4.09722	11.83140	1.000	-26.1019	34.2964
Native Group (Conv.)	Kanji Group	-1.95192	10.94135	1.000	-29.8793	25.9754
	Peer Group (Conv.)	-4.09722	11.83140	1.000	-34.2964	26.1019

Post-experiment Perception Survey

In this section, the results of the post-experiment perception survey will be reported. The data contained 31 students' responses. First, participants' overall reactions to conversation practice will be presented, followed by their beliefs about conversation partners. Except text entries and multiple choices (e.g., year in school, and sex), the survey deployed a 5-point Linkert scale (strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, strongly disagree). For the reports below, "strongly agree" and "somewhat agree" are combined and labeled "positive", "neither agree nor disagree" is labeled "neutral", and "somewhat disagree" and "strongly disagree" are combined and labeled "negative".

Overall reactions to the conversation practice

Table 17 presents the overall reactions of the participants to the conversation practice (Peer Group and Native Group). Overall, the survey found that the students in the Native Group had more positive experiences than the students in the Peer Group did.

Table 17. Overall reactions to the conversation practice

	Peer Group (n = 10)			Native Group (n = 8)		
Satisfaction with the Practice	**positive**	**neutral**	**negative**	**positive**	**neutral**	**negative**
I enjoyed the online speaking practice.	80 %	10 %	10 %	100 %	0 %	0 %
I think the online speaking practice helped me improve my Japanese skills.	80 %	20 %	0 %	100 %	0 %	0 %
Overall, I was satisfied with the online speaking practice.	70 %	30 %	0 %	100 %	0 %	0 %
If I have a chance, I would like the online speaking practice again.	60 %	40 %	0 %	100 %	0 %	0 %
Satisfaction with the Assignment						
The conversation topics for the online speaking practice were not very good.	10 %	20 %	70 %	0 %	25 %	75 %
The total number of the online speaking practice sessions throughout the semester was too many.	20 %	10 %	70 %	25 %	50 %	25 %
30 minutes for each session was too short.	0 %	30 %	70 %	12.5 %	37.5 %	50 %

Issues with conversation practice

The participants responded to the question "Were there any difficulties with the online speaking practice? If so, what?" by citing some issues. For example, a bad Internet connection was

a common problem that both groups experienced, and scheduling was another main difficulty for the Native Group. Other difficulties from the students' responses were as follows, with misspellings in the responses corrected in parentheses.

- I wouldn't call them difficulties as much as they are imperfections, but I realized that the conversation practice didn't push me into using new vocabluary (vocabulary) and grammar as I thought they should have. [a student from the Peer Group]
- I never really figured out how to start and end the conversation. [a student from the Native Group]

Some other comments

Ten participants commented in text for "Please tell us your opinion or comments regarding the online speaking practice if you have any." Seven comments out of the ten included positive reactions to the overall experience such as "it was a great/useful/beneficial experience/exercise". The other comments are listed below. For comments written in Japanese, translations by the present author follow in parentheses.

- 運良くパートナーが日本語が上手で楽しかったです。((I) was lucky to have a conversation partner that is good at Japanese, and (I) enjoyed it.) [a student from the Peer Group]
- 会話練習はいい練習でした、そして、それは授業に(で)とどまる(続けられる)べきです。(The conversation practice was good practice. And, it should stay in the course.) [a student from the Peer Group]
- I thought the speaking practice was really useful and beneficial, and I know it's tough, but perhaps matching those who have similar interests/speaking levels together to help both parties improve. [a student from the Peer Group]
- とてもいい練習でした。難しかったけど、私の自信を上げました。((It) was very good practice. (It) was difficult, but I became more confident.) [a student from the Native Group]

- I think it was great practice. It was the first and still the only chances I've had to talk to native people who are my own age. It was great to hear the phrases they use, their tones of voice, which modern words they have. It is the closest experience I have had to having true Japanese friends thus far. [a student from the Native Group]
- I think it's better when the conversation is not prepared but maybe students can prepare a bit of outline of what to speak. [a student from the Native Group]

The students' beliefs about conversation partners

The tables below present the students' responses about conversation partners. There were two questions related to the students' beliefs about conversation partners.

Table 18. Students' beliefs about conversation partners (Peer Group)

Peer Group (n = 10)			
	positive	**neutral**	**negative**
I was nervous speaking with my conversation partner.	30 %	20 %	50 %
I think I would have improved my Japanese more if I had had a native-speaker conversation partner.	50 %	40 %	10 %

Table 19. Students' beliefs about conversation partners (Native Group)

Native Group (n = 8)			
	positive	**neutral**	**negative**
I was nervous speaking with my conversation partners.	50 %	12.5 %	37.5 %
I think I would have improved my Japanese more if I had had a peer (classmate/coursemate) conversation partner.	25 %	0 %	75 %

For the Kanji Group, a similar question was asked in order to elicit their beliefs about a conversation partner. Table 20 shows the results.

Table 20. Students' beliefs about conversation partners (Kanji Group)

Kanji Group (n = 13)			
	positive	**neutral**	**negative**
If I had chosen the conversation practice, I would have preferred to have a native speaker as my conversation partner.	100 %	0 %	0 %

The survey found students' strong beliefs in native speaker status of their conversation partners, especially the results indicate that students who did not have conversation practice with a peer (Native and Kanji Group) tend to have stronger preference for a native speaker as their conversation partner.

The students' comments about conversation partners

The tables below list students' comments about their conversation partners. The question was "what were good or bad points of your conversation partner(s)?". Therefore, comments that do not directly respond to this question (e.g., scheduling difficulty) were excluded. Also, similar comments have been combined into one response. For comments written in Japanese, grammatical corrections and translations by the present author follow in parentheses.

Table 21. Students' comments about conversation partners (Peer Group)

Peer Group (n = 10)
• He is a really nice person. • I really enjoyed speaking with my conversation partner. He was already fluent in Japanese, so it was very valuable to listen and learn from them (him). It was a little difficult at first to get my ideas across to him in speech, but as we progressed(,) my fluency, accuracy, and speed grew. I am very grateful to them(him) for having such conversations with me. • 会話のパートナーは特に悪い点はなくて、スラスラと話が進める事ができるところがいい点だ(った)と思います。((My) conversation partner did not have bad points, and their good point was being able to smoothly carry on a conversation.) • We used translator occasionally. • We were both not that proficient in Japanese, which made it hard to communicate sometimes when both of us weren't sure what words to use. However, my conversation partner was very pleasant to work with and it was very fun to get to know them. • 僕や(の)質問の意味を理解していなかった((my conversation partner) did not understand the meanings of my questions.)

Table 22. Students' comments about conversation partners (Native Group)

Native Group (n = 8)
• They were all very kind. • They are Japanese. • わがる（かり）やすい (they were easy to understand.) • One good thing was that everyone was enthusiastic about the speaking practice. • Some of the people were a bit more or less sensitive to the fact that we are still learning Japanese I think. Some people I felt a little bad to say "I don't quite understand." But at the same time, almost everyone was understanding and kind and tried to speak a bit clearer if I was confused. It was hard to understand the speed of some people's speech, but that was ultimately really good practice. • Sometimes it took us a while to find out a good conversation topic but in general it was good. • 時々知らない日本語を使って、理解できませんでした。(They sometimes used Japanese that I don't know, so I couldn't understand it.)

Summary

This chapter presented results of students' improvement oral proficiency by the semester-long video-synchronous conversation practice, and the post-experimental perception surveys. The students' improvement was analyzed quantitatively. The results indicated that all the three groups improved the oral proficiency significantly, and the gains of each group were not significant. The results of the post-experimental surveys revealed that many of the participants in both groups had positive experience on the online practice. The surveys also revealed students' beliefs about conversation partners' native speaker status. Additionally, students' descriptive comments about the practice and the conversation partners were reported as well.

CHAPTER 5. DISCUSSIONS

This chapter will present discussions based on the results in the previous chapter and the research questions.

Research Questions #1 & #2

RQ1: Does weekly video-synchronous conversation practice with a conversation partner help L2 Japanese learners improve their oral proficiency? If so, to what extent?

RQ2: If weekly video-synchronous conversation practice helps improve L2 Japanese learners' oral proficiency, is there any difference between having a peer conversation partner and having a native conversation partner in improving L2 Japanese learners' oral proficiency?

First, improvement in oral proficiency of each group was analyzed using a paired t-test. The results indicated that all groups significantly improved their oral proficiency: the Kanji Group (M = 29.08, SD = 29.4), $t(12) = 4.13$, $p < .00$, the Peer Group (M = 31.22, SD = 24.61), $t(8) = 3.81$, $p = .01$, the Native Group (M = 27.13, SD = 22.12), $t(7) = 3.47$, $p = .01$. Then, a one-way ANOVA was run to see if there were any differences in their improvement among groups. The result revealed that the means of the gain scores of the three groups were not statistically different. From these results, it can be concluded that the participants of all the groups significantly improved their oral proficiency in the semester, and the effect of the weekly video-synchronous conversation practice in improving students' oral proficiency was not found in this study. This is because that the control group unexpectedly improved their oral proficiency as much as the experiment groups did.

The effect of the conversation practice was not clear for two reasons. First, the course contained a great number of speaking opportunities for students such as interview tests, individual video-making homework, an animation voice-over group project, an individual oral presentation, and reading-aloud homework as well as in-class verbal interactions and "Conversation Corner," an extra credit speaking activity. Indeed, according to the course syllabus, more than 45% of the

overall grade relied on speaking tasks; reading-aloud homework is not included in the percentage since it is a part of homework, and the syllabus did not state weight of each homework. Since these factors presumably solely caused the Kanji Group's statistically significant progress, they must have contributed to the Peer and Native Groups' improvement to a large extent as well. Hence, these factors made it difficult to isolate the effect of the online conversation practice on the improvement of the experimental groups.

Second, the students' attitudes toward the Japanese class seemed to differ to a high degree from individual to individual. According to the course instructor, the Kanji Group tended to have many enthusiastic students. For example, more students from the Kanji Group attended the Conversation Corner, properly completed homework, and more actively engaged in classes, than those from the Peer and Native Groups. Indeed, the instructor reported that two students from the Kanji Group attended the Conversation Corner almost all the time, one of whom appeared as an outlier with an outstanding amount of improvement. On the other hand, no students from the conversation groups attended it except one from the Native Group that sometimes did but not as often as the students from the Kanji Group. In contrast to the earnest students from the Kanji Group, there were some problematic students in the conversation groups; they did not always submit homework on time nor completed with adequate quality, and did not always actively participate in classes. Thus, although the course contents offered abundant opportunities to improve their oral proficiency, because of individuals' different attitudes toward learning, it is presumed that students' learning outcomes from the opportunities varied to a large extent.

This result does not support the findings of the previous studies (Kato et al., 2016; Saito & Akiyama, 2017) that reported significant differences in improvement between control and experimental groups with one exceptional result from the study of Kato et al (2016); they did not find significant differences in improvement between American control and experimental groups that were measured by mean utterance length. There are several possible causes for this discrepancy. Compared to the study of Kato et al (2016), the amount of practice (the total minutes of practice the target language) was smaller in the present research: 150 minutes fewer. In addition, the frequency of practice was less in this present research (once in a week) than theirs (twice a week). However, despite the larger amount of practice and the more frequent practice opportunity, Kato et al (2016) found that the improvements of the American experimental and control groups

that were measured by mean utterance length were not statistically different, and concluded that it was because of in-class instruction.

In comparison with the study of Saito and Akiyama (2017), even though the amount of practice in their research was less than that of the present research by 30 minutes, their data revealed that the experimental group significantly improved fluency while the control group did not. This inconsistency between these two studies could be because of the difference in participants' characteristics. The participants in Sato and Akiyama's study were Japanese learners of English inexperienced in speaking. Unlike them, students in the present study had had enough experience of speaking Japanese. For that reason, although the treatment protocols were very similar, the results turned out to be different.

Lastly, in terms of statistical analysis, the very small number of samples and great variations in the participants' oral proficiency scores within each group made it hard to produce significant results. All things above considered, to answer RQ1, we would have to say that the effect of weekly conversation practice was not found in this present study.

With regard to RQ2, all the groups including the Kanji Group significantly improved their oral proficiency, and the degree of improvement did not significantly differ among the three groups. Assuming that conversation practice accounts for a significant portion of improvement of Peer and Native Groups, no evidence was found in this experiment that either peer or native is better than the other as a conversation partner.

Research Question #3

RQ3: What are the L2 Japanese learners' reactions toward weekly video-synchronous conversation practice?

For this question, the conversation groups' responses to the seven questions ("satisfaction with the practice" and "satisfaction with the assignment") will be discussed below. The overall reactions of the participants toward video-synchronous conversation practice in the semester were positive. In terms of satisfaction with the practice, the Native Group showed a higher satisfaction level than the Peer Group; the average positive responses for the four related questions were 72.5 % (Peer Group) and 100 % (Native Group). There were two possible causes for the dissatisfaction of the Peer Group: mismatch of oral proficiency levels with their partner, and excessively difficult

conversation topics and conversation-generating questions. Even though the conversation topics and conversation-generating questions were adjusted to cover the textbook contents, they seemed to be too difficult for some peer pairs. As for the Native Group, the timing of using a native speaker as a conversation partner was a concern because the timing was a semester earlier than the suggestion of a previous study (Spring et al., 2019). Although some students commented that their conversation partners did not show consideration for the fact that they were still a learner, they concluded that it was still good practice. Overall, no negative feedback about this issue was reported on the survey, and fifth semester seemed to be fine for using a native speaker as a conversation partner with the students.

In terms of participants' satisfaction with the assignment, the number of sessions seemed to be adequate for the Peer Group, but not for the Native Group; 70% of the students from the Peer Group responded that ten times was not too many, but only 25% of the students from the Native Group responded the same way. This might have been because of the Native Group's extra work to contact their conversation partner and schedule a meeting every week. The duration of each session seemed to be adequate for both groups.

Research Questions #4 and #5

RQ4: What are the L2 Japanese learners' beliefs toward having a peer conversation partner in weekly video-synchronous conversation practice?

RQ5: What are the L2 Japanese learners' beliefs toward having a native conversation partner in weekly video-synchronous conversation practice?

In order to answer these questions, the conversation groups' responses to the question "I think I would have improved my Japanese more if I had had a native-speaker (Peer Group) / a peer (classmate/coursemate) conversation partner (Native Group)" will be discussed along with the related question "I was nervous speaking with my conversation partner." The two groups showed interesting preferences as to their ideal conversation partner for improving their oral proficiency after having had 10 conversation practice sessions; the Native Group showed a slightly higher preference for a native speaker as their conversation partner (75% prefer native speakers and 25% prefer a peer) than the Peer Group (50% prefer native speakers, 40% do not care, and 10% prefer

a peer). It seems that students' beliefs in native speaker partners in language learning remain strong even after satisfying semester-long experience of conversation practice with a peer. The Peer Group's experience of conversation practice might have resulted in the slightly weaker preference for native speakers, but it is unknown if their experience had changed their opinions since the survey was conducted only at the end of the semester.

Correspondingly, it is noteworthy that 10% and 25% of the students from the Peer and Native Group, respectively, preferred a peer as their conversation partner for improving their oral proficiency. These opinions are also understandable considering their nervous levels when speaking with their conversation partners; from students' responses to the question "I was nervous speaking with my conversation partner," 30% of the students from the Peer Group were nervous and 50% of them were not, while 50% of the students from the Native Group were and 37.5% of them were not. The slightly higher nervous levels of the Native Group correspond to the discussion of Varonis and Gass (1985); they found a larger number of negotiations of meaning in conversations between non-native speakers than in those between a non-native speaker and a native speaker, and concluded that was because non-native speakers were not threatened to show their incompetency to other non-native speakers. If we were to interpret this to mean that non-native speakers feel more at ease speaking with another non-native speaker than with a native speaker, it seems true that learners felt less threatened speaking with non-native speakers than with native speakers.

Research Question #6

RQ6: What are the L2 Japanese learners' beliefs about conversation partner?

For this question, the Kanji Group's responses to the question "If I had chosen the conversation practice, I would have preferred to have a native speaker as my conversation partner" will be discussed, assuming that they best represent L2 Japanese learners' belief in general among the three groups. Even though the question was not exactly the same as the ones for the conversation groups, their responses can be considered to reflect their beliefs about native speakers in language learning.

All (100%) of the students from the Kanji Group preferred a native speaker when given a choice, which differs from students' belief about in-class peer interaction in other studies (Abe,

2009; Matsumoto, 2020; Kishi, 2020). Based on these findings from the previous studies and the result of this study, it can be said that L2 Japanese learners believe that in-class peer interaction is helpful, but when they have a choice of native speakers, they prefer native speakers over peers. Assuming that the students from the Peer and Native Groups held the same belief before their participation in the conversation practice, it is possible that their experience with regular conversation practice changed their opinions.

Summary

This chapter discussed the results of the students' improvement of oral proficiency and the post-experimental perception surveys. The inconsistent results with the previous studies about the improvement were described by some possible reasons. Moreover, considerable reasons of the slightly different satisfaction levels of the practice by the two conversation groups were discussed. Further, students' beliefs in native speaker status were interpreted referring to students' responses from the three groups.

CHAPTER 6. CONCLUSION

This chapter will first present a brief summary of the findings of this study, and limitations, pedagogical implications, and future research directions will follow after it.

Study Findings

The present study investigated the effectiveness of weekly video-synchronous conversation practice in improving L2 Japanese learners' oral proficiency. The effect of the semester-long video-synchronous conversation practice was not found in the study. Because the control group also made significant gains comparable to the experimental groups, it was not possible to isolate the effect of the conversation practice. The most likely causes were the course content and the characteristics of the students in each group; the course placed a great deal of weight on the speaking skill, and the students in the Kanji Group were apparently more enthusiastic and motivated than those in the Peer and Native Groups.

In terms of students' reactions, they showed positive overall reactions to conversation practice. The satisfaction levels were examined in two ways: weekly video-synchronous practice as oral practice and as an assignment. As practice, the Native Group showed exceptionally high satisfaction levels although the Peer Group also showed positive reactions. As an assignment, the structure seemed fine for the Peer Group; ten times of practice was perceived as too many for the Native Group that had to arrange a meeting every week.

Furthermore, students' beliefs about conversation partners were investigated. Their responses demonstrated strong preferences for a native speaker as their conversation partners. The Kanji Group displayed the strongest preference for the native speaker status among the three groups followed by the Native Group, and the Peer Group. From these results, it can be said that L2 Japanese learners in the U.S. universities are inclined to believe that native speakers are more beneficial conversation partners than peers. Additionally, the results imply that positive experience of regular conversation practice with peers could change the beliefs.

Limitations

Since the experiment was conducted in an ordinary JFL university course, not in a laboratory, there were several uncontrollable parts. The discussions below will present five of the limitations of this study.

First, this study was not able to isolate the effect of the semester-long video-synchronous conversation practice because the course was designed with an extraordinary focus on the speaking skill, which ought to have influenced students' progress on oral proficiency a great deal. The present researcher was not able to control the course design because she was not the instructor in charge of the course.

Second, the participants' Japanese levels varied to a large extent including heritage speakers; native-like heritage speakers were counted as "learners" in this study. To the best of the present author's knowledge, there were two heritage speakers in the Peer Group, and they were matched with other peers. In such cases, their conversation partners in the Peer Group might have experienced conversation practice similar to that with a native speaker rather than that with a peer.

Third, the two experimental groups were not able to conduct the conversation sessions exactly in the same way in terms of conversation partners; the Peer Group had the same conversation partner every week while the Native Group did not. This was due to scheduling difficulty with the native speakers; students from the Native Group were also supposed to have the same conversation partner throughout the semester, but not all the native speakers were available during the entire period of the experiment.

Fourth, the amount of time in which students were engaged in conversation practice differed from student to student; although the requirement as homework was 30 minutes every week, some students conversed more than 30 minutes and the others did less than 30 minutes. Those who did not meet the requirement were told to follow the assignment guideline by the present author, but it was not perfectly controlled.

Lastly, the sample size was too small.

Pedagogical Implications

Although this study could not find conclusive empirical evidence for the effectiveness of the semester-long video-synchronous conversation practice, outside-a-classroom conversation practice can be a fun assignment for students; all the students (100%) from the Native Group and

60% from the Peer Group answered that they would like the practice again if they have a chance in the future. It might also raise students' confidence levels in speaking Japanese, which could affect their improvement in the long run.

If video-synchronous conversation practice is used as a part of coursework, careful consideration is needed in paring peers so that they maintain conversations all by themselves. In addition to conversation topics and conversation-generating questions for active interactions, it seemed important that the pairs share similar oral proficiency levels, personalities, and interests. Even though EI test scores can be a useful indicator of each student's oral proficiency level, it should be supplemented with other information when it comes to matching peers. This is because although EI is a valid oral proficiency test, students' communication skills cannot be determined solely by its score; maintaining active conversations or building rapport with their conversation partners needs other things like maturity and sociability along with oral proficiency. In this sense, teachers' subjective impression of students' speaking skill might help as an indicator of their communication skills. As for the matching of students' interests, like in the experiment of Kato, et al. (2016), it would be preferable to survey students' backgrounds and interests (e.g., study field, hobbies, and gender preference of their conversation partner) before they begin the practice. Combining these factors (EI scores, teachers' impression, and students' self-report survey) would produce better matches, which in turn would bring about more interactions during the practice.

Also, if conversation partners are to be fixed throughout a semester, checking in with individuals after a few sessions is advisable to prevent possible issues (e.g., difficulties of conversation topics, mismatch of oral proficiency levels, and personality conflicts). Moreover, when a smooth and active conversation is the main purpose of the activity, conversation-generating questions should be easier than the current level of students' knowledge, especially in the case of peer interaction. It would also be interesting to give students topics in advance, and let them prepare questions for conversation sessions.

Finally, when people from various cultures participate in the conversation (e.g., international students and native speakers from Japan in this study), brief instruction on polite and impolite topics of the native culture in which the institution is located and the target culture would be useful for participants to raise their cultural awareness.

Future Research Directions

In the present research, advised by the previous research of Spring et al. (2019), topics and conversation-generating questions were provided with students so they sustain conversations for 30 minutes. However, it needs to be examined whether open-ended opinion exchange is the best way to trigger active conversations or there are better activities such as information gap, decision-making or storytelling. From the perspective of Interaction Hypothesis (Long, 2015), the most effective conversation activities to improve students' oral proficiency should require a lot of interactions between the students. Thus, future research should investigate conversation activities that best fit video-synchronous practice.

Also, the focus of this research was to quantitatively measure the improvement of students' oral proficiency in semester-long video-synchronous conversation practice. In other words, analyzing the quality of the conversations (i.e. how they interacted with their conversation partners) was beyond the scope of this study. Future research might find an alternative explanation of the results of this study by analyzing their video-recorded conversations: e.g., checking and comparing the number of conversation breakdowns, and negotiations of meaning by interlocutor type.

APPENDIX A. EI SENTENCES

1. 今日は寒すぎるので、出かけない方がよさそうです。
2. 来週の授業はないらしいけど、明日はあるはずだよ。
3. 道が混んでいて、乗りたい電車に乗れなかった。
4. 熱がありそうなので、風邪を引いたかもしれません。
5. すぐ戻れるなら窓は開けたままでもいいよ。
6. 雨が降るかどうか分からないから、傘を持って行くことにします。
7. 午前 11 時なのにまだ寝ていたら、母に起こされた。
8. 先生がいらっしゃらないからもう帰ってしまってもいいかな。
9. 忘れやすい漢字は何回も書いたら、覚えられると思います。
10. 友達の誕生日にかわいらしい人形をあげたらとても喜んでくれた。
11. 今日の会議の時間を教えていただけませんか。
12. 日本人は働かせられすぎなんじゃないかと思う。
13. お酒は飲めるけど、あまり飲みすぎないようにしています。
14. 推薦状を書いてくださった先生にお礼を言いに行った。
15. 先週お目にかかった先生はインディアナ大学の先生だそうだ。
16. 大（だい）地震の時は、揺れる前に大きい音がするそうです。
17. 上司が部下に断れないお願いをするのはいじめと同じだ。
18. 生け花という伝統的な日本文化を習いに行くことにしました。
19. ゆっくり休んだ後は、しっかり働くことができるみたいです。
20. 何時間もかけて終えた宿題を寮の部屋に忘れてきてしまった。

APPENDIX B. POST-EXPERIMENTAL PERCEPTION SURVEYS

[ ] means types of answer formats. M=multiple-choice, L=Linkert-scale, and T=Text entries

Demographic-related Questions (for all three groups)

1. 何年生ですか。(What year in school are you?) [M]
2. 何歳ですか。(What is your age?) [T]
3. 性別は？(What is your sex?) [M]
4. 第一言語は何ですか？(What is your first language?) [T]
5. 専攻は何ですか？(What is your major?) [T]
6. 301 を取る前に日本語をどのくらい勉強しましたか。(Before taking the 301 course, how long had you studied Japanese?) [M + T]
7. 日本に住んだことがありますか。どのくらい住んでいましたか。(Have you ever lived in Japan? If so, how long have you been in Japan?) [M + T]
8. 日本に留学したことがありますか。どのくらい留学していましたか。(Have you ever studied abroad in Japan? If so, how long have you been in Japan?) [M + T]
9. 日本語のバックグランドがありますか。（ Do you have a Japanese language background? ）[M + T]

13 Reaction-related Questions (Peer and Native Groups)

1. どうして会話練習を選びましたか。(Why did you choose the conversation practice over the kanji storytelling?) [T]
2. オンライン会話練習は楽しかった。(I enjoyed the online speaking practice) [L]
3. オンライン会話練習の会話トピックはあまりよくなかった。(The conversation topics for the online speaking practice were not very good). [L]

4. オンライン会話練習は日本語の上達に役に立ったと思う。(I think the online speaking practice helped me improve my Japanese skills) [L]
5. オンライン会話練習で上達したと思う日本語のスキル。(My Japanese skill(s) that have been improved by the online speaking practice is/are) [M + T]
6. 学期を通してのオンライン会話練習の回数は多すぎた。(The total number of the online speaking practice sessions throughout the semester was too many.) [L]
7. 1 回 30 分のオンライン会話練習は短かすぎた。(30 minutes for each session was too short.) [L]
8. 会話のパートナーはよかった。(My conversation partner was good.) (Peer Group) [L] / 会話のパートナーの内何人よかったですか。(How many of your conversation partners were good?) (Native Group) [M]
9. 会話のパートナーのよかった点／悪かった点は何ですか。(What were good or bad points of your conversation partner?) [T]
10. 全体的にこのオンライン会話練習に満足した。(Overall, I was satisfied with the online speaking practice) [L]
11. 機会があったら、またオンラインでの会話練習をしたい。(If I have a chance, I would like the online speaking practice again.) [L]
12. オンラインでの会話練習をするにあたって、難しかったことがありましたか。何でしたか。(Were there any difficulties with the online speaking practice? If so, what?) [T]
13. その他コメントや感想があったら書いてください。(Please tell us your opinion or comments regarding the online speaking practice if you have any.) [T]

Belief-related Questions

Peer Group

1. 会話のパートナーと日本語で話すのは緊張した。(I was nervous speaking with my conversation partner.) [L]

2. ネイティブスピーカーの会話パートナーとオンラインでの会話練習をしたら、日本語がもっと上手になったと思う。(I think I would have improved my Japanese more if I had had a native-speaker conversation partner.) [L]
3. ネイティブスピーカーの会話パートナーとのオンライン会話練習の方が楽しめたと思う。(I think I would have enjoyed online speaking practice more with a native-speaker conversation partner.) [L]

Native Group

1. 会話のパートナーと日本語で話すのは緊張した。(I was nervous speaking with my conversation partners.) [L]
2. クラス／コースの学生が会話のパートナーだったら、日本語がもっと上手になったと思う。(I think I would have improved my Japanese more if I had had a peer(classmate/coursemate) conversation partner.) [L]
3. クラス／コースの学生が会話のパートナーだった方が楽しめたと思う。(I think I would have enjoyed online speaking practice more with a peer(classmate/coursemate) conversation partner.) [L]

Kanji Group

1. もし会話練習を選んでいたら、会話のパートナーはネイティブスピーカーがよかった。(If I had chosen the conversation practice, I would have preferred to have a native spear as my conversation partner.) [L]

REFERENCES

Abe, S. (2009). Japanese language learners' beliefs about language learning surveyed in a university in Madrid, Spain. *Journal of School of Foreign Languages, Nagoya University of Foreign Studies*, (37), 25-62.

American Council on the Teaching Foreign Languages. (n.d.). ACTFL/CAEP program standards for the preparation of foreign language teachers. ACTFL.

Bowden, H. W. (2016). Assessing second-language oral proficiency for research. *Studies in Second Language Acquisition*, 38(4), 647-675.

Chambless, K. S. (2012). Teachers' oral proficiency in the target language: Research on its role in language teaching and learning. *Foreign Language Annals*, 45(S1), S141-S162.

Fahmimroah, F., Ping, M. T., & Setiawan, I. (2017). English department students' beliefs towards plurilingualism and its influence on English academic achievement and language learning styles in Mulawarman University. *Script Journal*, 2(1), 51-67.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *The American Psychologist*, 34(10), 906-911.

Fukada, A. (2009) *Speak Everywhere*, licensed to and operated by e-Language Learning LLC, from

Fukada, A. (2013) An online oral practice/assessment platform: Speak Everywhere. *The IALLT Journal*, 43(1), 64-77. http://speak-everywhere.com

Fukada, A., & Wei. M. (2012) Nihongo kyooikugenbani okeru supiikingu shidoono ichizuketo genjoo-ankeeto choosani motozuite- (The current status of speaking training in Japanese language courses: A pilot survey study) In Matsumoto, K., S. Tomizawa, & G. Zheng (Eds.) Embracing the New Phase in Japanese Education in the United States (2013) 61-74.

Ginther, A. (2020). Lecture on Language Testing and Assessment. Purdue University, West Lafayette IN.

Heisig, J. W. (2011). Remembering the Kanji 1: A complete course on how not to forget the meaning and writing of Japanese characters. University of Hawai'i Press.

Horwitz, E. K. (1985). Using student beliefs about language learning and teaching in the foreign language methods course. *Foreign Language Annals*, 18(4), 333-343.

Horwitz, E. K. (1987). Surveying student beliefs about language learning. In Wenden, A., Rubin, J., Learner strategies in language learning (pp. 119–129). Prentice-Hall International.

Kato, F., Spring, R., & Mori, C. (2016). Mutually beneficial foreign language learning: Creating meaningful interactions through video-synchronous computer-mediated communication. *Foreign Language Annals, 49(2), 355-366*.

Kishi, T. (2020). Egiputojin nihongo gakusyuusyato egiputojin nihongo kyooshino biriifu -egiputo ainsyamusu daigakudeno choosa kara (The beliefs of Egyptian Japanese learners and Egyptian Japanese teachers: Survey at Ain Shams University in Egypt). *The Japan Foundation Japanese-Language Education Bulletin*, 16, 29-40.

Kissau, Scott. (2014). The impact of the Oral Proficiency Interview on one foreign language teacher education program. *Foreign Language Annals*, 47(3), 527-545.

Krashen, S. (1985). The input hypothesis: Issues and implications. London; New York: Longman.

Long, M. H. (1983*). Native speaker/non-native speaker conversation in the second language classroom*. ScholarSpace at University of Hawaii at Manoa.

Long, M. H., & Porter, P. A. (1985). Group work, interlanguage talk, and second language acquisition. *TESOL Quarterly*, 19(2), 207-228.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In Ritchie, W. C., Bhatia, T. K., Handbook Of Second Language Acquisition (pp. 413–468). essay, Academic Press.

Long, M. (2015). Second language acquisition and task-based language teaching (First ed.).

Matsumoto, M. (2020). Kosutarikajin nihongo gakusyusyano gengogakusyu biriifuno tokucho (The characteristics of language learning beliefs of costa rican japanese learners). *Saitama Gengo Kenkyu*, (4), 1-13.

Miller, A. M. (2019). Exploring achievement goal theory, ACTFL's 5 Cs, and the L2 classroom: What goals do students set? *Foreign Language Annals*, 52(2), 237-254.

Miller, J. F., & Chapman, R. S. (1975). Length variables in sentence imitation. *Language and Speech*, 18(1), 35-41.

National Standards in Foreign Language Education Project. (1996). *Standards for foreign language learning: Preparing for the 21st century*. Yonkers, NY: National Standards in Foreign Language Education Project.

Saito, K., & Akiyama, Y. (2017). Video-based interaction, negotiation for comprehensibility, and second language speech learning: A longitudinal study. *Language Learning*, 67(1), 43-74.

Schwartz, A. H., & Daly, D. A. (1976) Some explicit guidelines for constructing and scoring elicited imitation tasks. *Language, Speech, and Hearing Services in Schools,* 7, 33-44.

Spring, R., Kato, F., & Mori, C. (2019). Factors associated with improvement in oral fluency when using video-synchronous mediated communication with native speakers. *Foreign Language Annals*, 52(1), 87-100. https://doi.org/10.1111/flan.12381

Stansfield, C. W., & Kenyon, D. M. (1992). Research on the comparability of the oral proficiency interview and the simulated oral proficiency interview. *System (Linköping)*, 20(3), 347-364.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), Input in second language acquisition (pp.235-253), Rowley, MA: Newbury House.

Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough! The Canadian Modern Language Review, 50(1), 158-164. [7]

Terhune, N.M. (2016). Language learning going global: Linking teachers and learners via commercial Skype-based CMC. *Computer Assisted Language Learning*, 29(6), 1071-1089.

TrueNorth. (2018). *Technical report: Improving the TrueNorth speaking test predictive algorithm for oral language ability*.

Varonis, E. M., & Gass, S. (1985). Non-native/non-native conversations: A model for negotiation of meaning. *Applied Linguistics*, 6(1), 71-90.

Vinther, T. (2002). Elicited imitation: A brief overview. *International Journal of Applied Linguistics*, 12(1), 54-73.

Wang, J., Spencer, K., & Xing, M. (2009). Metacognitive beliefs and strategies in learning Chinese as a foreign language. *System (Linköping)*, 37(1), 46-56.

Wenden, A. L. (1999). An introduction to metacognitive knowledge and beliefs in language learning: Beyond the basics. *System (Linköping)*, 27(4), 435-441.

Yan, X., Maeda, Y., Lv, J., & Ginther, A. (2016). Elicited imitation as a measure of second language proficiency: A narrative review and meta-analysis. *Language Testing*, 33(4), 497-528.

Yasui, A., Ide, Y., Doi, M., & Hamada, H. (2019). *Quartet: Intermediate Japanese Across the Four Language Skills I*. (T. Sakamoto, Ed.). The Japan Times.