

# **App-based Language Instruction for Developing L2 Oral Proficiency in English**

**Leung, Ricky Chi Yan**

## **Introduction**

Over the past decade, mobile technologies such as smartphones, tablets, and other internet accessible electronic devices have become embedded into peoples' everyday lives all over the world. As of 2020, an estimated 67% (5.2 billion people) of the world's population had a mobile subscription to the internet (GSMA, 2021). The proliferation of mobile technology throughout the world has also led to an ever-increasing market for applications (apps) to operate on mobile technologies. The selection of apps for users has become highly personalized, reflecting how individuals chooses to use their time for social, professional, educational, or other personal purposes. In the domain of second language (L2) learning, various apps can be downloaded for prevalent operating system platforms such as iOS, Android, and Windows. One of the most popular language learning apps is Duolingo, which features over 400 million users with an average of 40 million active monthly users (Duolingo, 2021). Duolingo provides its users with affordances such as flexibility in time and place of use, accessibility of L2 lesson materials across multiple devices and platforms, and adaptability to an individual's usage approach. Although these affordances provide L2 learners with a platform designed for ease of use and potential L2 improvement, there has not been much empirical investigation into the purported benefits of such language learning apps. Plonsky and Ziegler (2016) identified that second language acquisition (SLA) research needs to be concerned with "how the affordances of technology might best be exploited to provide learners with optimal language learning opportunities" (p. 17).

## **Mobile Assisted Language Learning**

Early research on mobile assisted language learning (MALL) examined the L2 learning effects of distance education (Green, Collier, & Evans, 2001), the use of mobile phones for L2 instruction (Thornton & Houser, 2005), and the use of iPods as a voice recorder and L2 listening medium (Belanger, 2005).

Mobile technologies such as smartphones and tablets gradually began to emerge in the late 2000's and during the early 2010's, their presence had penetrated the domain of education and traditional classroom settings. By the mid 2010's, mobile technologies permeated learning environments with portable computers, tablets, and smartphones establishing new foundations for MALL. Language learners became endowed with tools and opportunities that were not available through traditional classrooms or face-to-face settings. Language learning websites, apps, and social networking tools provided instant accessibility to learning content anytime, anywhere. The increasing embedment of smartphones in daily life enabled people to start filling in pockets of "dead time" with the usage of apps, representing new learning dynamics regarding how people allocate their study time. Collaborative learning and social networking features also stimulated communication among app users by providing them with a forum to ask questions, find answers, and share their L2 learning process and experiences with others synchronously and asynchronously.

As MALL continues to expand in the field of SLA, the experimental results and L2 pedagogical outcomes of mobile technology use must be carefully examined. Existing research on popular commercial language learning apps such as Duolingo (Vesselinov & Grego, 2012), Rosetta Stone (Vesselinov, 2009), and Babbel (Vesselinov & Grego, 2016a), have stated favorable L2 learning outcomes for its users. These studies have also suggested that the L2 instructional content offered by these apps provide equal or greater effectiveness than face-to-face language courses (Loewen et al., 2019). Such claims need to be considered with caution as the Duolingo, Rosetta Stone, and Babbel studies were carried out by researchers affiliated to the company, thus demonstrating a high likelihood of bias in favor of such apps. In a meta-analysis of existing MALL studies, Burston (2015) stated that in 20 years of MALL project implementations, fundamental problems such as inadequate research design and an overly technocentric focus are commonly found in MALL studies. From 291 distinct MALL studies identified in his examination, Burston found that only 35 of the studies met minimal research design conditions of sample size and duration (a minimum of ten participants and a time period of at least one month). From the remaining 35 studies, Burston identified only 19 of the MALL studies to be reliable enough to serve as a basis for determining the learning outcomes of MALL applications. Of the 19 studies, 15 were considered to report positive L2 learning results in regard to reading and listening gains from using a MALL application. The remaining four studies with a focus on vocabulary, reported no significant differences (Burston, 2015). Albeit the supposed improvements in receptive skills such as reading, listening, and vocabulary learning, an essential skill that tends to be neglected in the L2 learning apps is speaking. In contrast to in-person instruction or computer assisted language learning (CALL) platforms, L2 learning apps do not offer users the functionality to practice or develop L2 speaking skills. This is a significant difference when compared to the video-based synchronous communication available on CALL platforms. Despite the lack of L2 speaking functionality, Lord (2016) suggested that language learning apps can be used as complements to traditional classroom instruction as they allow independent learners, who may lack access to traditional forms of L2 instruction, with a convenient, affordable and

purportedly effective means of developing L2 proficiency.

Although apps such as Duolingo, Rosetta Stone, and Babbel possess a large user base, the effectiveness of these commercial apps for L2 learning have received little attention from academic researchers. Loewen et al. (2020) suggested that limitations in researcher control when employing language apps as an experimental intervention act as a deterrence for researchers. Language learning apps possess an inherent degree of variability that the researchers cannot control for and therefore, this makes any empirical investigation into the effectiveness of app-based learning a challenging task. Another reason for the lack of academic research on app-based learning is the commercial nature of these apps. The companies and developers behind these apps tend to make audacious claims about rapid learning and being able to speak a foreign language in a limited amount of time (Loewen et al., 2020; Lord, 2016). With that being said, the current study presents an examination of the effectiveness of L2 learning for one of the popular commercial apps, Duolingo.

### **Duolingo**

Duolingo is an online language learning website and mobile-based app. The company was founded in 2011 by Louis Van Ahn and Henry Selinker, its platform launched in 2012, and in 2021, it became a publicly traded company, demonstrating its growth, popularity, and financial resources. Duolingo provides a number of affordances for its users, it is accessible worldwide through the internet, uses a freemium model (a pricing strategy in which Duolingo is provided for free, but money is charged for additional features that expand the functionality of the software, revenue is generated from ads), offers a continually growing number of languages and lessons to learn, incentivizes users with gamified features for fun and motivation, and offers users a single user account that can sync across multiple platforms and mobile devices across iOS, Android, and Windows operating systems. The company states, “Our free, bite-size lessons feel more like a game than a textbook, and that's by design: Learning is easier when you're having fun” (Duolingo, 2021).

Research regarding the L2 learning benefits accrued from the use of Duolingo are limited. Vesselinov and Grego (2012) stated that 34 hours spent learning Spanish on Duolingo produced similar results to taking the first semester of introductory Spanish in a college-level course. Vesselinov (2009) noted that in comparison to Duolingo, users of Rosetta Stone had to study between 55 to 60 hours to achieve similar results. Vesselinov & Grego (2012) further estimated that 34 hours using Duolingo is equivalent to the language instruction provided in a 16-week university semester course. Research on commercial language learning apps such as Rosetta Stone (Vesselinov, 2009), Duolingo (Loewen et al., 2019; Rachels & Rockinson-Szapkiw, 2018; Vesselinov & Grego, 2012), Babbel (Vesselinov & Grego, 2016a), and Busuu (Vesselinov & Grego, 2016b), have found positive L2 learning outcomes, particularly in terms of receptive knowledge of written L2 vocabulary and grammar. Krashen (2014) also noted that language learning apps such as Duolingo, reliably yield increases in learner vocabulary and grammar knowledge. Krashen further stated that the type of L2 knowledge gained from such apps is more explicit

and decontextualized linguistic knowledge, which can be accessed through reflection as time pressure is reduced during the use of such apps. Rachels and Rockinson-Szapkiw (2018) examined the use of Duolingo among elementary school students learning Spanish. Participants underwent either face-to-face instruction or Duolingo (40 minutes of exposure a week for 12 weeks). At the conclusion of the 12 weeks, Rachels and Rockinson-Szapkiw found that the two groups' performance on Spanish grammar and vocabulary tests were not statistically significant. Despite Duolingo's nonsignificant results in their study, Rachels and Rockinson-Szapkiw stated that the use of Duolingo presented users with an affordable and cost-effective option for L2 instruction. In a review of Duolingo, Cunningham (2015) commented that the gamification features on the platform give it an addictive edge for many learners to stay engaged. Munday's (2016) survey of beginner and advanced Spanish learners' perceptions of Duolingo also revealed that Duolingo's easy-to-use interface, accessibility, and gamified features made it an enjoyable app to practice Spanish. Falk and Gotz (2016) analyzed Duolingo's interactive and social networking features which included the use of discussion forums, following other users to share and compare progress results, and the use of a leaderboard function to compare progress results with a larger group of users. Through an online survey completed by 212 participants learning German (mostly university students and a few staff members) at two German universities, Falk and Gotz (2016) found that the majority of their participants found Duolingo's automated feedback, gamified and multimodal learning tasks, and overall use to be positive and helpful for L2 learning. However, only a small percentage of their participants indicated the use of interactive and social networking features available. For the participants that did use Duolingo's interactive features, the social aspect of being able to interact with other learners and being a part of a bigger "language community" was viewed positively (Falk & Gotz, 2016). Recently, Duolingo has added new interactional features such as events and podcasts. The effect of the "events" feature is particularly intriguing as language learners can meet individually or in groups online via Zoom (video conferencing app). This feature further pushes the conceptions of what a language community is and adds another dimension of online interaction, something of which might be worthy of future investigation.

In a mixed-methods study, Loewen et al. (2019) investigated the effectiveness of Duolingo in developing learners' L2 knowledge of Turkish. Using a Turkish 151 test (university course exam in Turkish) as the pretest and posttest measures of Turkish proficiency, the quantitative results from nine participants were varied. On the Turkish 151 posttest, the participants scored lowest on the speaking (33%) and listening (37%) sections but achieved higher scores on the reading (57%), writing (55%), and lexicogrammar (50%) sections. These results correspond to what Krashen (2014) stated, that apps like Duolingo promote greater gains of explicit knowledge rather than improve real-time L2 oral proficiency. After 34 hours of study on Duolingo, eight out of the nine participants received failing scores on the Turkish 151 posttest. These results call into question Vesselinov and Grego's (2012) claim regarding Duolingo's efficacy of being the equivalent of a university-level language course. Loewen et al. (2019) suggested a possible reason for the participants low test scores could be attributed to the

study of the target language, Turkish, as none of the participants had any prior knowledge of Turkish. The participants in Vesselinov and Grego's (2012) study were not complete beginners of L2 Spanish, possessing variable levels of proficiency, ranging from beginner to advanced. It is possible that L2 learners possessing a background or a previous level of proficiency in the target language might learn more L2 from the use of Duolingo than those who are complete beginners. Loewen et al. (2019) noted that all participants possessed more L2 Turkish knowledge at the conclusion of the study in comparison to when they began, citing the time on task factor. A moderate correlation between the amount of study time and test scores ( $r = .58$ ) was also identified. Additionally, the qualitative analysis revealed that several participants were not especially motivated by the choice of Turkish as the target language (Loewen et al., 2019) and this might have significantly influenced their L2 motivation. Moreover, the duration and frequency of study was found to be inconsistent among the participants with some participants using Duolingo on a consistent, daily basis for short periods (less than half an hour) of time whereas others choose to engage in intermittent, longer study sessions.

In a study on how language learning apps such as Duolingo shape the everyday experiences, motivations, and efficacy of L2 learning, Isbell et al. (2017) conducted a narrative analysis with three researcher student participants over a period of 12 weeks, studying the target language of Turkish. The participants recorded their thoughts in journal entries, maintained time logs of their Duolingo usage, and also took a Turkish 101 test (university course exam for introductory Turkish). At the end of the study, all three participants received failing scores on the Turkish 101 exam although Isbell et al. (2017) suggested that this might have been due to the mismatch the content on the Turkish 101 exam content with the learning content provided in the Duolingo course. In the participants' journals, each displayed fluctuations in their motivation at different stages of the study towards learning Turkish. The participants also found Duolingo's primary reliance on decontextualized grammar-translation exercises and audiolingual drilling as contributing factors to learner frustration and limited achievement, resulting in the participants' decision to discontinue learning Turkish, post study (Isbell et al., 2017). The participants also stated that the interactional feature of being able to follow friends, inspired healthy competition, which in turn resulted in completing more lessons for L2 practice. However, there was no opportunity to communicate orally in Turkish, resulting in the participants' self-questioning of their own L2 learning progress and proficiency (Isbell et al., 2017).

Academic researchers have called into question claims about the development of oral communicative abilities through app-based learning (Krashen, 2014; Lord, 2016). Despite their effectiveness in developing explicit, receptive knowledge of L2 reading, vocabulary, and grammar, the value of Duolingo for developing oral proficiency is unclear. For example, Lord (2016) found that while beginning-level Spanish learners had similar outcomes for classroom instruction and Rosetta Stone study, those who learned exclusively via Rosetta Stone struggled in conversation compared to learners who received classroom instruction. Loewen et al. (2019) identified that after a semester-long use of Duolingo, L2 learners' performance on oral test components was worse than performance on reading, writing, and

lexicogrammar test components in Turkish; a finding corroborated by the study's participants' self reports of difficulty in speaking Turkish. The current study addresses whether L2 oral proficiency can be developed through use of Duolingo with the following research questions:

RQ1. Can L2 oral proficiency be developed through consistent use of Duolingo in short durations of study sessions?

RQ2. What perceptions or experiences does a L2 English learner report after use of the app?

## **Methodology**

### **Participants**

One participant took part in this study. Yuko (pseudonym) is a native of Japan and her L1 background is Japanese. At the time of this study, her English level was self-reported to be “beginner” as she had taken required English classes from junior high school to high school in the Japanese education system. In university, she also enrolled in a few elective English classes out of interest but has never studied English beyond those experiences. In her early 20's, she spent one month in England for business purposes. That was the only lengthy period that she had ever spent overseas in an L2 (English) environment. Yuko is currently in her 40's and is a full-time working mother. She was selected for this quasi-experimental study since her everyday living situation is a realistic one for many working adults in Japan. Beyond working full-time, taking care of her kids, and attending to other daily tasks and responsibilities, she has little time to study an L2, despite her desire to do so. In fact, she mentioned that simply getting 15 minutes of distraction-free, concentration time to study is hard to come by. These reasons led her to start using Duolingo at the beginning of 2021, prior to this study, as its convenience and ease of accessibility allowed her to use the app and study, prior to going to bed at night. These factors fit within the study's research objectives and parameters and Yuko was asked and agreed to participate in this study. Following ethical research guidelines, informed consent was obtained from Yuko and all the information shared in this study regarding Yuko's personal details and statements were consented to on behalf of the participant.

## **Materials**

### **Duolingo's Educational Content**

Duolingo provides L2 learning targets in skills and each skill is divided into lessons. Depending on the target language, from beginning to end, an entire language course (or language tree) consists of variable numbers of skills and lessons. There are no set numbers of skills or lessons because they are consistently being added and updated by Duolingo's educational content developers, as well as by volunteer contributors. Learners from different L1 backgrounds who study the same or different L2s will have different language trees. For instance, a L1 English learner enrolled in a L2 Japanese course on

Duolingo may feature 40 skills in 160 lessons. A L1 Japanese learner enrolled in an L2 English course on Duolingo may feature 80 skills in 370 lessons. It depends on how each language tree is organized, developed, and evolves. Skills are presented to users in a linear order, with new skills becoming available as previous skills are completed. This linear progression can be seen in Figure 1:



Figure 1

*Duolingo Mobile App Skills Interface*

When a user selects a skill, the available number of lessons for that skill appears. Each lesson provides users with a list of words that are used in the lesson. For each lesson or skill, there is no contextual framework or meaning behind the lesson tasks. Once a user has completed a lesson or skill, a user can review or “strengthen skills” within the app to differentiate it from regular lessons. A symbol in each lesson indicates the “strength” of that skill up to a maximum level of five. When a user reaches level five, the symbol changes to a gold color. Activities are presented sequentially and the lesson “extends” itself if you get incorrect answers (Munday, 2016). If a user completes a lesson with no errors, it takes around seventeen to twenty language learning tasks to complete a lesson. Depending on the user, this is estimated to generally take five to fifteen minutes. Table 1 provides the types of L2 learning tasks available in each lesson on Duolingo and a short description of each task:

**Table 1**

*Duolingo's Lesson Tasks*

L2 Learning Task	Description
Vocabulary/picture matching	Write a vocabulary word after seeing a picture that represents it.
L2-L1 translations	Translate a sentence into your native language. When words are first presented, the user can hover over the word to see its meaning.
L1-L2 translations	Translate a sentence into the language being studied.
Dictation	Dictation: write a sentence that you hear. There are two speeds, normal and slow, which you can click to hear the sentence more clearly.
Oral repetition exercises	Pronouncing a sentence. Through voice recognition software, the app can detect whether your pronunciation is correct.
Vocabulary matching	Match pairs of words.
Sentence unscrambling	Put a series of scrambled words in order.
Multiple choice translations	Choose from three sentences in the target language to see which ones fit the sentence in the native language.

Interaction and social networking features allow Duolingo users to participate in forums, follow others' learning progress on leaderboards, and the option to join L2 learning-related events.

Duolingo also features gamified elements. Learners can earn experience points (XP) by completing lessons. A learner who sets and reaches daily XP goals is rewarded with "streaks", which represents a length of consecutive days of study. Completing skills, reaching daily goals, reaching the next XP level, or maintaining a streak are some of the ways to earn "gems" or "lingots", Duolingo's digital currencies. You can use gems or lingots for purchases in the app for bonus skills, power-ups, streak holders, and other aesthetic upgrades. These gamified elements are designed to make Duolingo more enjoyable and motivational, although the lesson tasks are traditional L2 learning exercises.

**Oral Proficiency Test**

In this study, a modified version of Payne and Whitney's (2002) Oral Production Interview Scale was used to measure the participant's English oral proficiency. Payne and Whitney's scale is measured on a 50-point scale (see Appendix A) and was designed to be sensitive enough to measure changes in oral proficiency that might occur in a single semester in a university course meeting only four hours per week. Although the participant in this study used Duolingo for less than four hours a week, Payne and Whitney's scale is more sensitive in measuring oral proficiency than a more prevalent and broad measure such as ACTFL's oral proficiency guidelines (see ACTFL, 2012). In the context of this study, L2 oral proficiency is defined as an individual's ability to produce L2 that is comprehensible with syntax and vocabulary appropriate to the task, is grammatically accurate, and is pronounced in a manner that approximates the speech of a very fluent non-native speaker (Payne & Whitney, 2002). For the oral



proficiency test, the participant selected one of three questions written in English (see Oral Proficiency Questions in Appendix B). The participant read the question and then spoke in English for approximately five minutes. If the participant had nothing else to say on a particular question, then the next question was chosen and the participant's answer given. The goal was to obtain a 5-minute speech sample from the participant. I administered the oral proficiency test twice, once for the pretest and once for the posttest. The participant's pretest and posttest scores were scored and rated by two L1 English raters who had no other involvement in the study. Both raters have over ten years of experience of teaching L2 English learners from a variety of cultural backgrounds at different age groups and education levels. The raters also possessed a degree of expertise in applied linguistics and SLA with one rater being an applied linguistics PhD student and the second rater having a PhD in applied linguistics. The pretest and posttest were scored by the raters. During the scoring process, the raters were instructed by the author to think of a very fluent nonnative speaker of English they know and consider that individual's language ability as a perfect score (the 50-point maximum). The two raters were sent a digital copy of the oral proficiency scale, audio files of the participant's pretest and posttest, and written instructions on how to use the scale. When the raters completed their scoring of the participant's pretest and posttest, they sent their oral proficiency scoring results back to the author. The author then compared the two rater's scoring for each oral proficiency test to maintain interrater reliability. For the oral proficiency pretest, interrater reliability on the 50-point scale was .92; on the posttest, interrater reliability was .88.

### **Participant Journal**

The participant maintained four brief journal entries for one month to record weekly reflections and personal observations of her learning experiences with Duolingo. The journal entries were recorded on text messages on the social networking service app, Line. This was done for convenience purposes and after each journal entry was complete, the information was sent from the participant to the author. Each journal entry was examined by the author and systematically coded for relevant themes and patterns, pertinent to the study.

### **Participant Interview**

At the conclusion of the study, a 25-minute, face-to-face interview was conducted with the participant. This interview was an opportunity to further gather more information and data regarding her perceptions and experiences of using Duolingo to learn English. The interview was recorded and analyzed by the author and relevant excerpts from the interview are provided later in the results section.

### **Procedure**

This study was conducted in four phases. In the first phase, prior to the commencement of this study, Yuko had already used Duolingo for 113 days consecutively. According to her estimate, she studied English on Duolingo for an average of five to fifteen minutes per evening, prior to going to bed.

This roughly equates to an average of 70 minutes per week. Beyond the use of Duolingo for L2 learning, Yuko did not engage in any other form of English language learning and relied exclusively on the app.

In the second phase of the study, the author conducted the oral proficiency pretest with Yuko in-person on June 6, 2021. After receiving the instructions for the oral proficiency test, Yuko answered three questions on the test and spoke into a voice recording app on the author's smartphone for five minutes. An audio file of Yuko's oral proficiency pretest was sent to the two raters and both raters completed their scoring by June 11, 2021, in which the scores were sent back to and compared by the author on the same date.

The third stage of the study took place between June 6, 2021 to August 2, 2021. Yuko's task was to continue using Duolingo daily to learn English. She was able to continue her daily streak of using Duolingo until the time period of July 26, 2021 to July 29, 2021, breaking her 170-day streak of Duolingo use. She continued her daily study on the app from July 30, 2021 to August 2, 2021.

In the fourth stage of the study, I conducted the oral proficiency posttest with Yuko in-person on August 3, 2021. The posttest procedures were the same as the pretest procedures. An audio file of Yuko's oral proficiency posttest was sent to the two raters and both raters completed their scoring by August 5, 2021, in which the scores were sent back to and compared by the author on the same date.

## Results

The current study addresses whether L2 oral proficiency can be developed through consistent use of Duolingo in short durations of study sessions and the reported perceptions or experiences of a L2 English learner after use of the app.

### Oral Proficiency Development Outcomes

Addressing the first research question, an analysis of Yuko's oral proficiency test scores was compared. Based on Yuko's pretest and posttest oral proficiency scores, it appears that there was no significant difference in a comparison of mean scores. The results of her oral proficiency test scores are shown in Table 2:

**Table 2**

*Yuko's Oral Proficiency Test Scores*

	Oral Proficiency Pretest Score (max = 50)	Oral Proficiency Posttest Score (max = 50)
Rater 1	25	24
Rater 2	27	27
Mean	26	25.5

On the pretest, Yuko got a mean score of 26 (interrater reliability = .92) and 25.5 for the posttest (interrater reliability = .88). A comparison of means shows that she did not improve her L2 oral proficiency, according to the 50-point scale of the oral proficiency test used in this study. According to her self-reported estimate, she had studied for an average of five to fifteen minutes per day for a total of 174 days. If her average daily study time of Duolingo was ten minutes per day, this would approximately equate to 29 hours of study. Her Duolingo profile (see Figure 2) showed a XP total of 9452 which is a high number, along with 102 “crowns”. In Duolingo, every time a user levels up a skill, they receive a crown. Duolingo does not provide any guidelines for XP or crown interpretations, so it is difficult to draw any sort of conclusions from this information.



Figure 2

*Yuko's Duolingo Profile*

### Perceptions and Experience of Using Duolingo

Addressing the second research question regarding the participant's perceptions and experiences of using Duolingo, an examination of Yuko's journal entries and interview responses was conducted. The data was analyzed, coded, and organized into three dominant themes: perceptions of using Duolingo, challenges and limitations of using Duolingo, and learner motivation using Duolingo. All original journal entries and interview excerpts from Yuko were in Japanese and have been translated into English by me for readability purposes. The journal and interview excerpt translations from Japanese to English were inspected for accuracy by a L1 Japanese and L2 English speaker. This individual (checker) is highly proficient in L2 English and is currently enrolled in a PhD program in applied linguistics conducted in English. Any errors in my original English to Japanese translations were corrected by the checker.

### **Perceptions of Using Duolingo**

Duolingo's accessibility and ease-of-use makes it an attractive language learning app for many learners. These features were mentioned in studies by Falk and Gotz (2016); Munday (2016); Rachels and Rockinson-Szapkiw (2018), and Yuko noted that this was one of Duolingo's benefits in her first journal entry:

Duolingo is easy to use. There is nothing difficult to use about it. It is also fun, has a nice design, and the characters are kind of cute as well. (Entry 1 excerpt).

Yuko viewed the aesthetic design and operability of Duolingo to be positive. In her journal entries, she did not mention anything regarding Duolingo's interactive features unlike the participants in other studies (Falk & Gotz, 2016; Loewen et al, 2020) that viewed Duolingo's interactive features as a language learning affordance. The topic of Duolingo's interactive features was discussed by Yuko during the interview:

[Interview, 8/3/2021]

[15:22]

Yuko: I know that there are some features to follow others and see their progress scores, but I don't have time for such extra things. There is also some kind of chat forum (referring to the discussion forum), but again, I don't have time. This is the only way. It might not be the most effective but even if I can get in 5 minutes before bed, that is better than nothing.

How she allocated her study time played a major role in her choice of what Duolingo features to use. The participants in other studies (Falk & Gotz, 2016; Munday, 2016) were primarily university students. Such research participants tend to have more time and resources available to engage in L2 study in comparison to Yuko who is working full-time employment and has parental responsibilities.

Yuko also commented on Duolingo's language and lesson content in the following entry:

New words come up from time to time in the lessons and I learn them. But I know most of them. For grammar, I know most of it. Just reviewing. It is mostly things that I have come across before so I don't feel any progress in that area. Reading: I can read English better than before. Listening: I can listen better than before. Speaking: No effect. I need to have many conversations in English. (Entry 1 excerpt).

Despite being a beginner English learner, Yuko wrote that she already knew or had previously come across the vocabulary and grammatical content featured in Duolingo's lessons for Japanese learners of English. She noted improvements in her L2 reading and listening abilities, but did not clarify in detail. Previous research (Krashen, 2014) has suggested that the use of language learning apps results in

improvements in receptive skills such as L2 reading, vocabulary, and grammar. Yuko further commented in the interview about her perceived self-improvement in L2 reading:

[Interview, 8/3/2021]

[8:54]

Yuko: I felt that my reading has improved which was surprising. I usually get the newspaper and in the newspaper on Saturday, there is an English section. Recently, when I look at that section, I can naturally interpret the reading. Rather than getting the feeling of “surprise”, I feel as though I can generally understand the reading. In the past, I would usually translate it but recently when I read it, I don’t necessary read it with Japanese translation and understand it in English.

Yuko felt “surprised” as she was able to read the English section of the newspaper. This might suggest that her reading comprehension skills improved with the use of Duolingo. Although improving her L2 reading skills was not her primary objective in using the app, she appeared to be pleased with development in this area. Yuko also discussed her improvement in L2 listening ability:

[Interview, 8/3/2021]

[10:12]

Yuko: What I have improved is my listening. I am more able to connect the words and the pronunciation. Before when the pronunciation was too fast, I couldn’t catch all of the words, but now, I can clearly hear all of the words individually. I couldn’t understand native pronunciation but now, I feel as though I can understand more English pronunciation. However, for my own pronunciation and repetition, no one is evaluating me so I feel as though I don’t know if my pronunciation is good.

Yuko stated that being able to comprehend the native-like English pronunciation of individual vocabulary was an improvement for her. Duolingo’s library of native-like English speech and pronunciation in dictation and oral repetition exercises in the lessons might have contributed to her L2 listening improvement. Although the dictation and oral repetition tasks used in Duolingo are not new pedagogical methods, their multimodal functionality in which users can touch individual words or sentences on the app screen and hear them repeated for parsing and further processing is an important learning aid for L2 learners.

### **Challenges and Limitations of Using Duolingo**

Prior to the commencement of this study, Yuko stated that one of her main goals was to improve her English speaking. She was hoping that by completing the lessons and tasks in Duolingo, that would in turn lead to improvement in her English oral proficiency. Her thoughts regarding her progress in English speaking were written in this journal excerpt:

I can now understand English without translating it into Japanese. But I still can't speak it out loud. It is important to have an environment (training) to speak English. It's hard to master English with just this. I need to study with other textbooks and talk with people. (Entry 4 excerpt).

Yuko's comment highlights one of the biggest weaknesses of mobile app-based learning is the lack of L2 speaking development. Including a dedicated speaking feature on language learning apps is a difficult task and is most likely being worked on by Duolingo developers. Yuko did not have high expectations that her speaking ability would develop using Duolingo and this sentiment was stated by Loewen et al. (2020) in that learners' expectations from app-based learning should be realistic. Loewen et al. further commented that learners might develop some speaking ability, but most of the learning will be in the areas of explicit grammar and vocabulary knowledge.

During the interview, Yuko added more thoughts about developing her English-speaking skill on Duolingo:

[Interview, 8/3/2021]

[17:55]

Yuko: There is no chance to interact or talk so that is one of the limitations. It is limited to a machine and not a human so I miss that human aspect. When I went to England for a month for work, it was a really fun time and when I talked to the locals, even at such a low level of English at the time, I thought that if I stayed there for a year or two, I would be able to speak English in a year or two. I was younger then and now I can only do about 5 minutes a day, so I do realize that it is very difficult to improve my English.

She emphasizes her awareness of Duolingo's limitations and expresses her lack of time to study as the primary obstacle in her English improvement. Moreover, she talked about an earlier period in her life and the possibility that she could have become more proficient in English had she chosen to stay in England. There is not much existing SLA research on "time periods" of acquisition for working adults due to moderating variables such as limited time resources, employment, parental responsibilities, and other factors that influence L2 learning. This theme was expressed in further detail by Yuko in her journal entry:

I also don't have time to take a lesson which I obviously know is better and would like to do, but if I did take lessons, the time I need to take care of my own things, talk to my kids, take care of housework; my schedule would get too overwhelming. In the morning, I also have about 30 minutes that I could dedicate to learning English, but again, if I use that time, my schedule will get too tight so at a minimum this is the best that I can do. That's a good point for me though because it is better than doing absolutely nothing! (Entry 3 excerpt)

### **Learner Motivation Using Duolingo**

From the onset and throughout the study while recording journal entries, Yuko did not express fluctuations in her motivation to learn L2; she simply consistently used the app daily which might reflect her actions as motivation. This contrasted with findings in Isbell et al. (2017) and Loewen et al. (2020) whose participants experienced a general decrease in motivation the longer they used Duolingo. Yuko described some of her feelings regarding motivation in her following journal entry:

Everyday I get the reminder notifications from Duolingo and it reminds to keep working hard and keep trying. When I get the notifications, I think, "Oh, I better start and study tonight." (Entry 2 excerpt).

One of the gamified features in Duolingo is "streaks". Users can maintain their streak as long as they consistently use the app for an extended period of time. Yuko brought up the topic of her "streak" during the interview:

[Interview, 8/3/2021]

[3:02]

Yuko: I did it for 170 days in a row and if you don't use it on a daily basis, your streak will reset. I didn't want it to break my streak. Last week I was too busy so my streak went back to three days. I was a disappointed because I tried hard to keep my streak going.

For Yuko, Duolingo's daily reminder notifications were seen as motivational, but for participants in Isbell et al. (2017), they were deemed as a nuisance. Maintaining her streak was important and this brings forth the question of whether she was using Duolingo to simply maintain her streak or did the streak facilitate L2 learning and serve its purpose as a reminder not to stop? Other Duolingo users share their streaks on leaderboards or on forums, but for Yuko, it was just for herself, suggesting high levels of intrinsic motivation to maintain the streak or learn L2 as a means of maintaining the streak.

Yuko further expressed her thoughts on motivation in the following interview excerpt:

[Interview, 8/3/2021]

[21:20]

Yuko: When you complete a lesson, you can do it again and it gets harder. That's interesting for me. But basically, I would say that my motivation has not really increased. It's impossible to get to a high native level, but I would like to be able to speak like an 8-10-year-old in English, and to be able to use it to communicate with people from other countries while travelling. Also, it is like my hobby, English is interesting and I like learning so that is why I persist.

Yuko's love of learning, desire to use L2 for traveling, and obtaining a level of basic English oral proficiency akin to a young L1 English native speaking child are her primary objectives. As motivation is difficult to measure and subject to available time periods and resources, it would be interesting to see if apps such as Duolingo facilitate any of these motivations into real-world outcomes over longer periods (over one year) of study and use.

## **Discussion**

The study results suggest that interpretations of L2 oral proficiency development through language learning apps such as Duolingo is minimal when daily usage durations range from five to fifteen minutes of study per day. Learners' expectations of L2 oral proficiency development should be tempered in contrast to claims made in previous studies (Vesselinov, 2009; Vesselinov & Grego, 2012; Vesselinov & Grego, 2016a, Vesselinov & Grego, 2016b) of purported gains in L2 oral proficiency from the usage of language learning apps. Akin to findings in (Isbell et al., 2017; Loewen et al. 2019, Loewen et al. 2020), the current study's participant reported improvements in receptive skills such as explicit reading, listening, and vocabulary. These findings align with theories in SLA that feature a facilitative role in explicit instruction (Krashen, 2014).

Time on task was another key factor that facilitated the study participant's improvement in receptive L2 skills. Yuko used Duolingo for 174 days in total, for a duration of over 28 hours on a highly consistent basis. Studies by Falk and Gotz (2016), Loewen et al. (2019), and Loewen et al. (2020), also identified time on task as a key predictor for the improvement of L2 explicit skills using language learning apps. It is important to keep in mind that although it is easy and convenient to start L2 learning on these apps, it can also be easy to stop or quit as well. Therefore, long-term studies (over a period of one year) on app-based learning are needed to determine the effects on L2 development.

Yuko perceived Duolingo's accessibility as one of the key factors in her usage of the app. By being able to access the app and study at times which fit her schedule, this was one of the main factors for her consistent, long-term use. Participants from existing studies (Falk & Gotz, 2016, Munday, 2016, Loewen et al., 2019) echoed similar sentiments as Duolingo's accessibility, convenience, and flexibility made it attractive for learners due to personalized nature; each individual can customize their learning schedule, L2 learning goals, participate in the Duolingo "community" with discussion forums or following other



members. Yuko was able to customize her L2 learning sessions according to her busy schedule, even sparingly at five to ten minute durations, to continue her L2 learning progression. Duolingo's gamified features also attempt to instill a sense of motivation through streaks, XP, and gems or lingots. Such gamified features make the app fun to use, like a game or competition, and would be a good addition to traditional pedagogical classrooms (Munday, 2016). Rachels & Rockinson-Spakiw (2018) also stated that gamification apps may be able to provide scaffolding for L2 learning in addition to traditional classroom instruction. As stated by Yuko, maintaining her streak was important and she did not want to break up her consistency in using the app. Therefore, the app's gamified features might have created a motivational flow within her L2 learning in response the challenge of maintaining her streak as well as gaining XP daily to meet her goals.

The limitations of Duolingo and app-based language learning are displayed by a primary reliance on decontextualized grammar-translation activities and audiolingual drilling. These are extensions of traditional pedagogical approaches to L2 learning and their ease of implementation, structure, and controllability of outcomes (correct or incorrect answers) are probably the primary reasons such tasks make up the language learning content featured on Duolingo and other mobile apps. These lesson activities and tasks might result in limited learner development and learner frustration, leading to the lack of long-term persistence in using Duolingo (Isbell et al., 2017; Rachels & Rockinson-Spakiw, 2018). Along with spaced repetition, these traditional L2 pedagogical methods are efficient, but lack contextual meaning and interactive communication with others. The next phase of app-based language learning needs to incorporate lessons or pedagogical materials which can facilitate L2 speaking opportunities in either a synchronous or asynchronous fashion. Beyond developing the receptive skills, productive skills need to be engaged in a manner that leads to oral proficiency development. How to incorporate L2 oral proficiency development facets into an app such as Duolingo has yet to be determined and is a work in progress. Perhaps machine learning virtual assistants such as Apple's Siri might be the next step forward. L2 learners can engage with a "chatbot" like Siri for oral production practice. This would seem ideal as the convenience of being able to speak an L2 and receive real-time feedback would be an enormous development not only for MALL, but for SLA in general. This technology does not currently exist, but it is not impossible as natural language generation, processing, and machine learning continues to improve in performance over time. In ten years' time, it might be possible to use an app such as Duolingo to talk and interact with a virtual interlocutor for oral proficiency development.

### **Conclusion**

Although the popularity of language learning apps such as Duolingo have been increasing, empirical evidence regarding L2 learner outcomes because of using such apps is relatively scarce. In the domain of app based, L2 oral proficiency, there are even fewer studies on learners' L2 speaking development outcomes. This study attempted to unveil if L2 oral proficiency development was possible

with Duolingo, albeit various limitations. Due to this study's small sample size of one participant, there is limited generalizability to the greater population. It is recommended that this study be replicated with larger sample sizes with other adult L2 learners in a context similar to Yuko's for results that possess greater statistical power to show the effects of the intervention. Another limitation is the lack quantitative measures for the participant's L2 output for the pretest and posttest. A detailed measurement, calculation, and analysis of specific lexis, syntax, or other linguistic structures produced by learners might identify what linguistic units are acquired and produced using L2 learning apps. The short data collection period in this study also did not allow for longitudinal effects to be measured on what would happen with a prolonged and continued use of Duolingo.

Duolingo, as currently stands, is a widely used L2 learning app that contains traditional L2 pedagogical lessons and content delivered in a modern, gamified experience to engage and motivated learners. To improve app-based learning, more recent advancements in SLA pedagogy might be integrated into these apps with a particular emphasis on creating learner opportunities for L2 speech. Despite overstated claims from Duolingo and other popular apps such as Rosetta Stone, and Babbel, existing research and the results from this study suggest that these platforms can help learners develop receptive skills and explicit knowledge in reading, vocabulary, grammar, and listening. Regarding the development of L2 oral proficiency with such apps, this has so far been found to be inconclusive.

### References

- American Council on the Teaching of Foreign Languages. (2012). *ACTFL proficiency guidelines 2012*.  
<https://www.actfl.org/sites/default/files/guidelines/ACTFLProficiencyGuidelines2012.pdf>
- Belanger, Y. (2005). *Duke university iPod first year experience final evaluation report*.  
[https://web.archive.org/web/20070609084848/http://cit.duke.edu/pdf/ipod\\_initiative\\_04\\_05.pdf](https://web.archive.org/web/20070609084848/http://cit.duke.edu/pdf/ipod_initiative_04_05.pdf)
- Burston, J. (2015). Twenty years of MALL project implementation: A meta-analysis of learning outcomes. *ReCALL*, 27(1), 4-20.  
<https://doi.org/10.1017/S0958344014000159>
- Cunningham, K. J. (2015) Duolingo. *TESL-EJ*, 19(1), 1-9.
- Duolingo. (2021). *Duolingo is now a public company!*  
<https://blog.duolingo.com/duolingo-ipo>
- Falk, S. & Gotz, S. (2016). *Interactivity in language learning applications: A case study based on Duolingo* (T. Zeyer, S. Stuhlmann, R. D. Jones, Eds.). *Interaktivität beim Fremdsprachenlehren und -lernen mit digitalen Medien: Hit oder hype?* Narr Francke Attempto Verlag GmbH + Co.
- Green, B. A., Collier, K. J., and Evans, N. (2001) *Teaching tomorrow's class today: English by telephone and computer from Hawaii to Tonga* (L. E. Henrichsen, Ed.). Teachers of English to Speakers of Other Languages, Inc.

- GSMA. (2021). *The mobile economy 2021*.  
[https://www.gsma.com/mobileeconomy/wp-content/uploads/2021/07/GSMA\\_MobileEconomy2021\\_3.pdf](https://www.gsma.com/mobileeconomy/wp-content/uploads/2021/07/GSMA_MobileEconomy2021_3.pdf)
- Isbell, D. R., Rawal, H., Oh, R., & Loewen, S. (2017). Narrative perspectives on self-directed foreign language learning in a computer- and mobile-assisted language learning context. *Languages*, 2(4), 1-20.  
<https://doi.org/10.3390/languages2020004>
- Krashen, S. (2014). Does Duolingo “trump” university-level language learning? *The International Journal of Foreign Language Teaching*, 13-15.
- Loewen, S., Crowther, D., Isbell, D. R., Kim, K. M., Maloney, J., Miller, Z. F., & Rawal, H. (2019). Mobile-assisted language learning: A Duolingo case study. *ReCALL*, 31(3), 293-311.  
<https://doi.org/10.1017/S0958344019000065>
- Loewen, S., Isbell, D. R., Sporn, Z. (2020). The effectiveness of app-based language instruction for developing receptive linguistic knowledge and oral communicative ability. *Foreign Language Annals*, 53(2), 1-25.  
<https://doi.org/10.1111/flan.12454>
- Lord, G. (2016). Rosetta stone for language learning. *IALLT Journal of Language Learning Technologies*, 46(1), 1-35.  
<https://doi.org/10.17161/iallt.v46i1.8552>
- Munday, P. (2016). The case for using Duolingo as part of the language classroom experience. *RIED*, 19(1), 83-101.
- Payne, J. S. & Whitney, P. J. (2002). Developing L2 oral proficiency through synchronous CMC: Output, working memory, and interlanguage development. *CALICO Journal*, 20(1), 7-32.  
<https://doi.org/10.1558/cj.v20i1.7-32>
- Plonsky, L., & Ziegler, N. (2016). The CALL-SLA interface: Insights from a second-order synthesis. *Language Learning & Technology*, 20(2), 17-37. <http://dx.doi.org/10125/44459>
- Rachels, J. R., & Rockinson-Szapkiw, A. J. (2018). The effects of a mobile gamification app on elementary students’ Spanish achievement and self-efficacy. *Computer Assisted Language Learning*, 31(1-2), 72-89. <https://doi.org/10.1080/09588221.2017.1382536>
- Thornton, P., & Houser, C. (2005). Using mobile phones in English education in Japan. *Journal of Computer Assisted Learning*, 21(3), 217-228.  
<https://doi.org/10.1111/j.1365-2729.2005.00129.x>
- Vesselinov, R. (2009). *Measuring the effectiveness of Rosetta Stone*.  
[http://resources.rosettastone.com/CDN/us/pdfs/Measuring\\_the\\_Effectiveness\\_RS-5.pdf](http://resources.rosettastone.com/CDN/us/pdfs/Measuring_the_Effectiveness_RS-5.pdf)
- Vesselinov, R., & Grego, J. (2012). *The Duolingo efficacy study*.  
[http://static.duolingo.com/s3/DuolingoReport\\_Final.pdf](http://static.duolingo.com/s3/DuolingoReport_Final.pdf)
- Vesselinov, R., & Grego, J. (2016a). *The Babbel efficacy study*.  
<https://press.babbel.com/en/releases/downloads/Babbel-Efficacy-Study.pdf>
- Vesselinov, R., & Grego, J. (2016b). *The busuu efficacy study*.  
[https://blog.busuu.com/wp-content/uploads/2016/05/The\\_busuu\\_Study2016.pdf](https://blog.busuu.com/wp-content/uploads/2016/05/The_busuu_Study2016.pdf)

## Appendices

### Appendix A

Oral Proficiency Rating Scale, adapted from (Payne & Whitney, 2002)

#### Comprehensibility

Score	Ability Description
10-9	For a native speaking interlocutor, it is possible to understand the nonnative speaking interlocutor without any confusion or difficulty.
8-6	For a native speaking interlocutor, it is possible to understand the nonnative speaking interlocutor with minimal difficulty.
5-3	For a native speaking interlocutor, it is possible to understand the nonnative speaking interlocutor with some difficulty.
2-1	For a native speaking interlocutor, it is possible to understand the nonnative speaking interlocutor with great difficulty.

#### Fluency

Score	Ability Description
10-9	Speaker demonstrates native-like fluency and uses hesitations only when appropriate.
8-6	Speaker demonstrates near native-like fluency and uses few hesitations or pauses.
6-5	Speaker demonstrates fairly continuous speech including some hesitations or pauses.
4-3	Speaker demonstrates speech that is disjointed including frequent hesitations or pauses.
2-1	Speaker demonstrates speech that is very disjointed including many hesitations or pauses.

#### Vocabulary Usage

Score	Ability Description
10-9	Very extensive vocabulary usage.
8-6	Good vocabulary usage, very few inappropriate terms.
6-5	Moderate vocabulary, a few inappropriate terms.
4-3	Limited vocabulary, some inappropriate terms used.
2-1	Very limited vocabulary, frequent use of inappropriate terms.

### Syntax and Grammar

Score	Ability Description
10-9	Native-like grammar and syntax; used a variety of syntax and tenses.
8-6	Near-native grammar and syntax; few mistakes.
6-5	Used few syntax structures, some grammar and syntax mistakes.
4-3	Very limited in syntax and grammar usage with frequent mistakes.
2-1	No systematic use of grammar and syntax rules.

### Pronunciation

Score	Ability Description
10-9	Native-like pronunciation, virtually no discernable accent, no errors.
8-6	Near-native pronunciation, slight accent, few errors.
6-5	Some errors; obvious accent, but doesn't interfere with comprehension
4-3	Frequent errors; strong accent; some comprehension difficulties.
2-1	Little effort to use English pronunciation; comprehension impeded.

### Appendix B

#### Oral Proficiency Questions

##### Pretest

Question 1: Tell me about your most recent trip.

Question 2: Tell me about your plans for Obon?

Question 3: Tell me what you do in a normal week?

##### Posttest

Question 1: Tell me about your most recent trip.

Question 2: Tell me about your plans for Obon?

Question 3: Tell me what you do in a normal week?

