

DUOLINGO AND HELLOTALK: ANALYZING AND EVALUATING LANGUAGE
LEARNING APPLICATIONS USING SLA THEORY AND USER FEEDBACK

BY

ROBERT CHARLES SCHROEDER

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master's in Arts in Teaching English as a Second Language
in the Graduate College of the
University of Illinois Urbana-Champaign, 2021

Urbana, Illinois

Master's Committee:

Associate Professor Dr. Randall Sadler, chair

Visiting Professor Dr. Lisa Pierce

ABSTRACT

This study performs an evaluation of two language learning apps using established SLA principles and user-feedback through a survey and interviews. A survey asking users about level appropriateness, focus on form, Zone of Proximal Development and scaffolding, interaction, communicative language, noticing, and motivation and gamification was administered and followed up by interviews for details. Results showed that users generally rated Duolingo well in most capacities with its strongest factor being vocabulary for specific contexts and scenarios, but that the app was rarely rated as excellent in any category because of oversimplified grammar instruction, poor focus on form in terms of pronunciation, and some aspects of the app that lead to user frustration. HelloTalk however was often rated evenly across categories and ratings or in some cases had users give opposite or scattered ratings, with the general defining factor in satisfaction being whether the user found a reliable and helpful language exchange. Implications of these findings for app developers and app users is discussed in the conclusion.

CONTENTS

Chapter 1: Introduction	1
Chapter 2: Literature Review	4
1) MALL	4
2) CALL.....	10
3) SLA	12
4) Research Questions and Hypothesis.....	20
Chapter 4: Methods.....	22
1) App review and guiding questions.	23
2) Survey Content.....	26
Demographics.....	27
Frequency and Duration of Use.....	27
Macro Questions.....	28
Level Appropriateness & Comprehensible Input and Output.....	30
Focus on Form (FoF)	31
ZPD and Scaffolding	32
Interaction	33
Communicative Language	33
Noticing.....	34
Motivation and Gamification.....	35
3) Survey Participants	36
4) Focus Groups and Interviews	40
Chapter 5: Analysis & Discussion.....	43
1) Frequency & Duration of Use.....	45
2) Duolingo Macro Questions: What’s Important For Learners and Overall Effectiveness.....	47
3) Level Appropriateness & Comprehensible Input and Output	51
4) Focus on Form.....	59
5) ZPD and Scaffolding	65
6) Interaction.....	66
7) Communicative Language	68
8) Noticing.....	73
9) Motivation and Gamification.....	75
Chapter 6: Conclusion.....	81
1) Summary of findings.....	81
2) Implications for App Developers.....	83

3) Implications for App Users.....	86
4) Limitations of This Study.....	87
5) Suggestions for future research.....	89
References	91
Appendix A: Survey.....	96
Appendix B: Duolingo Question 16 responses	127
Appendix C: HelloTalk Question 11 responses	128
Appendix D: Duolingo question 23 responses.....	129
Appendix E: HelloTalk Question 18 positive responses.....	130
Appendix F: HelloTalk question 18 negative responses	131
Appendix G: HelloTalk question 23 positive responses	132
Appendix H: HelloTalk question 23 negative responses.....	133
Appendix I: Sarah quote.....	134

CHAPTER 1: INTRODUCTION

The use of smartphones has exploded since the first iPhone was introduced in 2007. In the United States alone, roughly 85% of Americans say they own a smartphone of some kind according to a PEW research poll in February of 2021. This number gets even higher depending on the country, with advanced economies generally having higher smartphone penetration than developing economies. In comparison to the US, South Korea has a very high smartphone ownership rate of 95%, according to a similar international PEW poll that was conducted earlier in 2018. The PEW research center found that even emerging economies often have significant rates of ownership, up to 60% in countries like Brazil and South Africa. The high rate of smartphone ownership worldwide has many implications for communication, information proliferation, and education. While there is a stereotype of the smartphone addicted teenager walking into water fountains and traffic, the reality is that these smartphones are being used for things other than social media. In addition to the proliferation polls, the PEW research center (2014) also conducted a survey to find the common attitudes towards, and uses of, cell phones by most of their users in the US and found that these devices are being increasingly used for information gathering, online banking, job information and applications, and educational content. Overall, 30% of users used their smartphones for educational content, with that number increasing depending on age group with 44% of ages 18-29 using their smartphones for educational content. With the proliferation of these devices and the availability of free educational apps and programs the smartphone may be capable of revolutionizing education in general, especially for language learning, as the availability of education becomes more widely available to those with lower incomes.

Living in South Korea for a time, I noticed the frequency with which learners used their cell phones to help themselves during the language learning process. It was not uncommon for language exchanges to have their phone translator ready as a quick dictionary when a word was not understood, or for friends of mine to regularly use language learning applications like Duolingo to practice the local language so that they could communicate on a basic level. In Korea especially, the high rate of cell phone ownership becomes apparent the moment you step onto the subway; rows of heads are down staring at the glowing screens of their phones and tablets, and barely anyone would be looking up. Occasionally, I would catch a glimpse of Duolingo on someone's phone or notice the Rosetta Stone logo out of the corner of my eye as well as many of the popular Korean English learning apps. While some people might scoff at the fact that everyone is sitting staring at their phones, it is clear that a large portion of people, if not most, are using these smartphones to make the mundane moments of their life more productive. These experiences made me curious about the number of people using these applications for language learning, what the teaching potential of these applications are and, for those who are using them regularly, in what capacity they are relying on these applications for their language learning experiences.

Language learning apps come in many shapes and forms, but one of the most popular and recognizable ones is Duolingo, which boasts over 500 million users worldwide, with English being the most popular language learned using the app. The application is free and makes money by selling translation software that is fueled directly by responses on the app (Garcia, 2012) as well as advertisements in the app itself. While it commodifies the learning experiences of its users, this model makes language learning accessible to anyone who possesses a smartphone or a computer that can access the website. A more detailed breakdown of the program will be

discussed as part of this research, but the general background of Duolingo is that it is a rote-learning program that delivers sentences that must be translated to or from the target language by either typing in the answer or creating the sentence from a word bank. HelloTalk, by comparison, is a much newer application, currently claiming over 25 million members. The philosophy of the application is also significantly different, as the focus of the app is that users should “learn by chatting”, providing tools for translation, correction, and assistance to be used by language partners at their discretion. Both applications are free with purchasable extras, making most of the program available to anyone with a smartphone.

The potential of free mobile language learning applications, especially in terms of accessibility, is very promising. However, the field of Mobile-Assisted Language Learning (MALL) is relatively new, and information regarding the usefulness and learning outcomes of users of these applications is very sparse. Additionally, the guiding principles and instructional philosophies behind the pedagogies and materials of these applications are opaque in some cases, which begs the question as to what theories, models, research, or philosophies inform the developers design, and whether there is evidence that theories and practices from SLA have been used.

The objective of this study is to investigate these two applications for influence of SLA principles focusing on concepts such as level appropriateness, focus on form, scaffolding, use of interaction, communicative language, noticing, and gamification. The other objective is to investigate how learners use the apps and their perceptions about them.

CHAPTER 2: LITERATURE REVIEW

The literature that was reviewed for this study was varied but can be split into three camps. First, there are articles related specifically to Mobile Assisted Language Learning and its effectiveness. While the research on language learning apps is very limited, there is still some information available that can provide insight on how effective or useful apps might be for regular users. The second type of research that was reviewed was literature on Computer Assisted Language Learning (CALL) because the field has been around for a little longer and because CALL methods are often closely related to mobile learning methods. Many of the insights provided by data from CALL studies can be extrapolated to a MALL context, as many of the apps and programs are similar while smartphones have the added benefit of portability. Finally, a review of SLA theory and many of the established theoretical principles that have been shown to be effective in classroom settings will be covered. The end of the literature review will contain an overview of the research questions.

1) MALL

Mobile Assisted Language Learning (MALL) is the study of research on language learning and teaching through cellphones and smartphones, which can include use of applications or basic phone functions like text-messaging or calls. In terms of research, MALL is relatively new and lacks many conclusive studies. The research shown below covers a lot of the challenges related to MALL projects and the lack of solid evidence for the effectiveness of MALL studies, as well as MALL studies that give some insight into how these apps can be useful.

The potential of language learning apps and MALL has not gone unnoticed by researchers. In 2011 Robert-Godwin Jones (2011) commented on the new phenomenon at the time and its potential to help with language learning, especially with the rise of language learning

apps for specific languages and flashcard applications for learning vocabulary. Additionally, Jones suggested that with the capability of communicating with others in real-time, and the existence of built-in GPS, smartphones have the potential of facilitating meaningful interaction with one's environment and with others in the target language. The idea of meaningful interaction is particularly important to this study, as the apps have the potential of giving students useful, real-life language for travel or in HelloTalk's case, real social interaction with peers.

Jones came back to the topic of smartphones in 2017 and lamented the state of MALL studies, stating that many studies were experimental in nature, short term, extremely narrow, and often lacked many of the supports needed for definitive results. A good example of these limitations is Wang, Hwang, Yin, and Ma (2020) who tested students' ability to learn vocabulary from a MALL project with participants from Chinese university classes. This study showed only a slight, statistically insignificant advantage in favor of students who used the MALL program for their learning and consisted of a sample size of twenty-eight students. Despite limitations in sample size and the inconclusive results, MALL still has potential, and Jones cited Stockwell in arguing that many CALL-based projects, a field of study that has had more time for development, transfer over to MALL as well. More than just transferring, Jones also explains that MALL has additional benefits including global connectivity and personal empowerment because of customization features and availability.

This disappointment in the lack of solid research on the subject was repeated by Jack Burston, who wrote an article named *The Reality of Mall: Still on the Fringes* (2014), mentioning that even though MALL has existed for decades at that point and had a variety of options for use on a large scale in the form of apps, most studies on the subject were still extremely limited. However, even from the limited data he was able to acquire, Burston was able

to ascertain that smartphones were becoming a huge asset to language learning students and were being implemented primarily for vocabulary acquisition. It still stands though, through this article and another meta-analysis of the subject by Burston (2014) that the field of MALL still lacked studies that provide any objective, quantifiable data. This lack of objective available data makes it difficult to draw any conclusive assumptions about the topic and about the apps in question without further research, giving credence to the fact that these applications and uses need to be studied in further detail if we are to understand their implications for language learning.

While Burston's articles may suggest that he is confident about the usefulness and applicability of MALL, he also mentions that many current MALL projects have severe limitations. In another 2014 study, Burston found that most MALL projects and research have implemented MALL in teacher-centered, drill-based projects. However, while the applications might be useful for vocabulary retention as hinted by the aforementioned Wang, Hwang, Yin, and Ma (2020) study, Burston argues that there is a myriad of technical and situational constraints on implementation. Additionally, Burston argues that there are technical challenges to offering useful feedback to students as well. Implementation is not impossible though, as mobile phones offer an opportunity for real-time collaboration, personalized instruction based on student needs, and increased accessibility. As an ever-evolving app, recent upgrades to Duolingo's program may achieve some of these, and HelloTalk is specifically designed around the idea of real-time communication. It is also still possible that many of these applications may have some of the same shortcomings that Burston was concerned with. However, as long as studies continue to be limited in scope, most of the conclusions by these studies are just food-for-thought.

A commonly found side effect of these limitations is that many studies with similar goals tend to end in contradictory results. For example, Saricoban and Özturan (2013), through a study involving sending core vocabulary in the form of text messages to students between classes, found that the use of smartphones proved helpful for the students who used them in their study regimens, contradicting the statistically insignificant advantage shown by Wang, Hwang, Yin, and Ma (2020). While this study targets vocabulary acquisition, the possibilities for the portability and general technology use in a classroom had broad implications for the use of apps as supplementary material and allowed students to access their vocabulary whenever and however they wanted. However, this study is plagued by the exact same limitations and eludes any definitive results or conclusions about the benefits of MALL implementation.

A final, and huge, obstacle to MALL and language learning applications and research that should be mentioned is the issue of motivation. Al-Sabbagh, Bradley, and Bartram (2019) find that a usable interface for an application is vital to help with student motivation. In their study, they found that migrants in Sweden often did not use the available apps due to a lack of user-friendliness. There are some studies that explore other motivational variables as well; a small-scale study on motivation and language learning apps by Garcia Botero and Questier (2016) found that many students who dropped Duolingo did so because of social obligations and a lack of external motivation and a thesis study done by Tuncay (2020) mentions that one contributing factor to app attrition, or the rate at which users stop using the application, is that personal obligations often got in the way of learning. This issue of motivation is important to this study, as intrinsic and extrinsic motivation is one of the SLA principles that will be reviewed when investigating these apps.

Additionally, the issue of motivation might be a determining factor in the language learning outcomes of the learners who are using these programs. Nielsen (2011) found that students using programs for self-study were more likely to suspend their learning from students taking classes and other more resource-intensive methods such as in-person classes and paid tutoring. This is consistent with findings by Tuncay (2020) who found similar results suggesting that students who were involved in classes were more personally invested. These studies suggest that accessibility might not be everything, as the external motivations of the students will likely affect whether they keep using it on a regular basis.

Additionally, Kathia Ibacache (2019) conducted a survey among librarians to determine the proliferation of mobile language learning apps and the librarians' opinions regarding those applications. The data revealed that ninety-six out of 199 respondents had used mobile apps, but many of them did not think that the applications met their professional needs in their library work for communication, cataloging, and instruction. These doubts over the applications' efficacy for professional use is something a lot of learners might struggle with when using an app. There is some evidence that an additional external factor may be a learner's familiarity with technology that might affect their overall attitude towards a learning program, as Stiller and Koster (2016) found in their study that the most common factor for students dropping out of online forms of learning were student attitudes towards the technology itself. While these results suffer the same limitations as many of the others mentioned here, that there are likely external factors and motivations that will affect the motivation of students to continue learning a language.

From Burston's initial optimism towards the utility of language learning apps, there are plenty of studies and analyses that show the promising nature of applications and MALL outside

of his recommendations. One such study discussing the efficacy of applications for language learning was done by Kang and Lin (2019). While this study mainly focuses on the use of applications being used in China, and the analysis is not very detailed, the commentary made about the usefulness of these applications, especially for the fast-paced learning styles of working adults, also speaks to the potential strengths of language learning applications in the west as well. Specifically, Kang and Lin (2019) discuss applications that are intended for vocabulary memorization, listening, and speaking exercises and the potential benefits of these apps to learners as supplementary material. The vocabulary aspect of this is corroborated by Walker (2015) who conducted a study to find the usefulness of Memrise for long-term memorization of words and found that students who regularly integrated the program into their study regimen had an advantage over students who did not. Additionally, Kang and Lin's commentary on the proliferation of smartphones and learning apps and the implications of their widespread usage finds many parallels in the west as well, especially when considering the use of smartphones in younger generations. This suggests that the widespread use, availability, and portability of language learning applications may be their biggest potential strength.

There are some very recent, potentially positive, articles that focus on one of the applications being studied, with one of the more interesting ones being a case study on Duolingo done by Shawn Loewen (et al. 2019) who found that the application was promising as educational material, but that there was a wide variation in results depending on student motivation and frustration. These results were found through having a series of students keep journals and take regular assessments while learning from Duolingo, from which statistics and data could be analyzed. While the participants did show language improvement overall, Loewen (et al. 2019) admits that the research might be skewed by the fact that all participants were

students that were familiar with SLA theory. The frustration mentioned here might echo the previously mentioned familiarity and user-friendliness issues that cause learners to stop using the applications in question. The most promising aspect of this article regarding the proposed study is that, for the students who did utilize the application throughout, there was improvement. This establishes the potential for the utility of motivational materials, because a well-designed app that is capable of keeping its learners motivated may be a huge boon to language learning communities they serve.

Finally, one of the most relevant MALL articles to this proposed study so far is by Read and Barcena (2020) who have proposed a general framework for mobile assisted language learning. The framework here considers educational principles as well as some SLA principles. Some of the relevant guidelines for their framework include promoting plurilingualism, linguistic diversity, addressing communicative language competence, contextualizing the language process, and addressing different situations that students can find themselves in. Read and Barcena are not the only authors who have suggested general frameworks for evaluating language learning apps as well, Rosell Aguilar (2016) also suggested an evaluation framework based around language learning concepts, pedagogy, user experience, and technology. While incomplete, frameworks like these are a great start and provides credence to the idea that, with some guidelines, mobile language learning has great potential and can be evaluated based on specific educational standards. This framework is also important for establishing the idea that applications can be evaluated based on similar principles.

2) CALL

As smartphones and language learning applications are a relatively new phenomenon, it is helpful to fall back on some useful articles concerning CALL for insight on how technology

can influence the learning experience. However, it is important to remember that even CALL research is developing slowly while programs and applications are created at breakneck speed. In a 2009 article, Chapelle discussed the issue of SLA theory being outpaced by practical instruction resulting in a disconnection between the two. In this article, Chapelle focuses on how CALL is making this disconnect even wider, which has serious implications for modern disconnect between SLA theory and MALL.

The first relevant article that should be mentioned about this issue is from Sauro (2009), who investigated what types of computer-mediated feedback were most effective. While the statistics that came from the study showed no significant differences between the two types of feedback provided, metalinguistic feedback and recasts, both groups did show some improvement. However, the study is unclear on why the differences were not very pronounced, and the sample group and overall study scope are both limited as well. Even though there wasn't a major *difference* in learning outcomes between students receiving computer-mediated feedback, steady improvement from both sample groups shows that learning via computer-mediated feedback is possible and practical on some level, and more recent studies such as a meta-analysis done by Chang and Hung who claim in their abstract that their “. . . findings provide concrete evidence in favor of technology use. . .” in relation to language learning and feedback (2019) show that there is potential here. The feedback left on a computer and on a smartphone are likely to look very similar, and because this research suggests that feedback given through technology has the potential to be useful if the types of feedback that work best via technological mediums can be ascertained in some way, it also implies potential in smartphone applications as well.

There are also studies that suggest CALL is effective for vocabulary learning similarly to vocabulary studies for MALL. Hirschel and Fritz (2013) found that students performed moderately better on a vocabulary test after using a CALL based program to study. An additional study by Allum adds to this by specifically pointing out the potential of CALL for teaching new vocabulary. Both of these studies mirror some of the earlier MALL studies that showed MALL to be useful for vocabulary learning, and these results can be extrapolated to MALL implementation as well.

3) SLA

Articles on SLA principles and their implementation in classrooms are plentiful. For the purposes of this study, articles were found that directly relate to the principles that will be looked for in the applications. While SLA theory is still ongoing and the best practices and instructional methods are constantly in flux, the applications in question must get their inspiration for the app design from somewhere. However, the guiding philosophies of these applications are only hinted at rather than explicitly stated, so an investigation will have to be completed to find evidence of these philosophies in the apps themselves. While not all SLA theory is applicable to implementation within a language learning application on the phone, there are a few that can realistically be applied, and thus looked for in this investigation.

First of these is Focus on Form. It is accepted that for higher proficiency levels some type of form-based instruction on the target language is needed. Hummel (2014) claims that there “. . . is increasing recognition that L2 learners benefit from some type of explicit attention to aspects of linguistic form”. Other authors in the field have also pointed out studies that showcase the usefulness of some sort of instruction on form as well. This is supported by one particularly notable study by Williams that showed that lower-proficiency learners in the studied group did

not notice many of their own errors without explicit demonstrations of their errors through a series of memory and translation tasks (1999). This suggests that the applications will need some form of well-executed focus on form to be valuable in some way to the learners.

There is some evidence that the type of feedback when reviewing forms matters as well. In a 2011 study, Chang found that the grammar translation method was more effective in teaching grammatical structures than communicative language teaching. While the study was small in scope, and Chang does not claim that grammar translation is a useful tool for comprehensibility and communicative competence, this still suggests that certain methods of teaching form may be useful for certain purposes.

There are also studies that suggest student preferences and ability might affect their reception of feedback. In an overview of research done at the University of Victoria, Chen and Nassanji (2018) found that while most types of feedback worked on some level, most students preferred implicit, general feedback on repeated errors instead of explicit feedback on a single one-time error. An additional study by Yilmaz (2012) that examined the effectiveness of implicit and explicit feedback with learners of different working memory capacity found that explicit feedback worked well with students with both high working memory and low working memory, while implicit feedback did not work well with learners with low working memory. With these studies in mind, it is possible that, in a program that relies very heavily on an algorithm for its feedback, constant corrections on trivial errors may contribute to learner frustration and lack of motivation, and due to reliance on an algorithm the amount of explicit feedback may be limited.

Additionally, the frequency of corrections might have an impact on student frustration and motivation as well. Doughty and Varela (1998) investigated the effects of communicative focus on form and found that some students might have limits for their tolerance to form-focused

feedback. Something interesting brought up in Doughty and Varela's (1998) article that might have implications for language-learning apps is the evidence that some students were not comfortable with being corrected more than once or twice at a time, potentially worsening frustration with instructional materials, considering that a computer program cannot gauge student preferences or read a student's mood.

While it cannot be expected for applications such as Duolingo to utilize all feedback types, especially without the presence of an in-person teacher, there are a few feedback methods that automated technology can provide. Explicit feedback, explained by Lightbown and Spada (2017) as the ". explicit provision of the correct form" is something that applications could realistically provide through automated feedback. Implicit Feedback, such as recasts which involves ". . . reformulation of some or all of the student's utterance, minus the error." (Lightbown and Spada, 2017) could also be expected from automated systems. In both cases of explicit and implicit feedback Lightbown and Spada were referencing in-class interactions. While the feedback would look different in an online format and would be likely less focused on verbal communication and more on writing short sentences, the feedback can still either explicitly explain grammatical errors in some form or possibly just show the correct form afterwards.

Another element of SLA theory that will be explored in relation to these apps is learner motivation. Motivation has already been discussed a little bit in this review under MALL, and learner motivation is a principal factor in the outcome of language learning. Hummel (2014) notes that "Motivation, in particular, is commonly cited as playing a major role in any language learning situation. . .", noting a series of studies by psychologist Robert Gardner as principal evidence for this relationship. Two types of motivation noted by Hummel are Intrinsic

motivation, where a student has an internal, personal motivation for learning a language, and extrinsic motivation which is generally characterized by the expectation of some sort of award. Likewise, the motivation of a student using an application is likely to have similar effects on the student's outlook of that application, especially in their opinion of how successful the application was.

More importantly, applications can implement extrinsic motivation. The use of extrinsic motivation strategies in classroom settings has been shown to influence student outcomes, most notably through a study done by Guilloteaux and Dörnyei (2008) who observed the motivation strategies of teachers and identified a correlation between those strategies and positive engagement of the students in class. Language learning applications may have methods for maintaining similar interest and positive outcomes in their users.

Gamification is a currently well-known method for motivating students that applications can utilize. Gamification is a well-known and studied phenomenon in education circles. Kevin Werback, in his 2014 article explains gamification as making the learning process more fun, interesting, or incorporating rewards systems and other mechanisms to influence behavior in a certain way like videogames. In terms of language learning applications, the desired behavior would be continued use of the application, but by extension would include better learning outcomes.

Duolingo is particularly known for its gamification of language learning. Rachels & Szapkiw (2018) discuss the potential of Duolingo's gamification of learning by equating it with "flow theory", which posits that that a high level of satisfaction can be gained from a difficult activity if it includes some motivational elements, some of which include, but are not limited to, meaningful goals and clear feedback. In their study, they found that Duolingo's gamification of

language learning had a positive effect on elementary school students' motivation, and a positive effect on learning outcomes. However, no studies are available that suggest that this same gamification works on adults, and some of the studies in the MALL section, such as the one by Nielson (2011), have suggested that the motivation of adults might not be as easy to maintain or may be intrinsic in nature.

In this study, motivation and methods for maintaining it will be investigated for the applications. For this section, a study done by Noels, Pelletier, Clement, and Vallerant (2000) is useful because it handily establishes the relationship between types of motivation and outcomes and how different motivational levels affect performance and perception of production. A basic overview of their findings is that personally motivating factors such as travel, knowledge, and friendship generally resulted in a student having high internal motivation as well. Meanwhile there was an anomaly in the research that suggested that external motivation can have separate factors from internal motivation, such as puzzle solving or treating the learning experience as a game. Exact motivations and concrete studies on motivation are hard to pin down, but the results of this study are significant for the review of language applications. Particularly, the idea that gamification might motivate some learners while others might only be reached through internal motivation. An investigation into how these applications motivate their students will delve into some of these principles.

Something that is often overlooked in motivational studies is the effect of difficulty on student motivation. Gass and Selinker (2008) find through a review of various studies that motivation tends to be affected greatly by the perceived difficulty of an activity. Basically, if an activity has a moderate difficulty, students are more likely to be motivated to complete it. However, activities that are perceived as too easy or too difficult might have the opposite effect.

This is important to keep in mind for language learning applications, as starting every student at a beginning level might discourage experienced language learners from using the applications. Additionally, if activities are too difficult, it is possible that beginner users will just not like the app.

Input might also influence motivation as well as outcomes. Hence, comprehensible input is something that each of these applications will also be evaluated for. Stephen Krashen (2008) Discusses in detail the use of the Comprehension Hypothesis, which argues that students learn from what they can comprehend. While there are a series of studies discussed by Ortega (2013) that cast doubt on the idea that comprehensible input is the only thing one needs to learn a language, this proposed study will still look at input as it was designed for the learner. Specifically, the principle of comprehensible input will still be useful in determining if the materials provided are appropriate for the skill level being taught.

The comprehensible input hypothesis is not ironclad. In addition to Ortega, Hummel (2014) points out that Krashen received a significant amount of criticism for his attempts to provide an equation that determined comprehensible input. Additionally, Comprehension and understanding does not exactly mean that the language is being absorbed. To adjust for this reality, Gass and Selinker (2008) separate the concepts of comprehensible input and intake, where input is understood for functional purposes while intake is incorporated into the student's linguistic ability. Measuring the input that language learning applications give learners will be possible, however, measuring the intake of students would be significantly more difficult for this study.

Related to comprehensible input, Scaffolding and the Zone of Proximal Development (ZPD) are concepts that are well studied in educational circles. Vygotsky (1978) Introduced the

ZPD as the gap between what a student can do on their own and what they cannot realistically do, with the zone in-between being where learning occurs. A student would require the help of a teacher to complete activities in this area of the ZPD. By extension, Vygotsky (1978) also pioneered the concept of scaffolding, as the ZPD concept implied a connection with what a learner has already learned and what they are currently learning. Thus, applications and classrooms ideally wish to build-up the knowledge base in their students so they can complete more complicated tasks, something that language learning applications will need to incorporate to be effective.

Another important principle that will be looked at, especially when delving into HelloTalk, is interaction. Gass and Selinker (2008) define interaction as “. . . learning through input (exposure to language), production of language (output), and feedback that comes as a result of interaction. Lightbown and Spada (2017) expand on this by mentioning that the concept of the ZPD is changing a little bit, in that it previously referred to the experience between a teacher and a learner of lower ability in the target skill, current research suggests that this experience can also be between two people on the same skill level who are collaborating to complete an activity. This concept is directly related with negotiation of meaning, which “. . . refers to instances in conversation when participants need to interrupt the flow of conversation in order for both parties to understand what the conversation is about. . . “ (Gass and Selinker, 2008). The concept of interaction is something that directly ties to the HelloTalk model of language learning, which specifically facilitates learning through communication with peers, one of the main points of investigation being whether the interactions on HelloTalk offer something that Duolingo cannot, interaction and collaboration, and whether or not students are able to get valuable learning experiences from that.

There are some studies that corroborate the concept of interaction for learning. Loewen and Sato (2018) Discuss interaction and instruction, especially the relationship between theory and classroom instruction. However, the most pertinent part of the article in relation to interaction is their discussion on peer interaction and its shown positive effects on pronunciation, negotiation of meaning, and form. This could prove to be a very strong aspect for HelloTalk, as the tools they provide not only help with focus on form but assist with pronunciation and communication as well.

A final aspect of SLA to be discussed is noticing. Mackey (2006) found that there is a positive relationship between students noticing their errors and their learning gains in the classroom. While noticing is related to the feedback that is provided to students, Kartchava and Ammar (2014) found that noticeability is often related to the specific forms being focused on, with students noticing errors in some forms better than others. This may have implications for the applications' abilities to help students notice their errors and suggests potential barriers to noticeability outside of application design.

There is one study that is worth mentioning that does not directly connect to any of the above reviews of literature but should be discussed because it strived to achieve something similar to what is being tried here. A thesis written by Petterson (2019) is one of the few studies to be found that attempts an analysis of the use of SLA principles in language learning apps. In this study, language learning applications were evaluated for their implementation of proven SLA instructional theories and found that implementation was scattered among the apps. While this study is not really a peer-reviewed academic article, and it does not incorporate user perspectives in its analysis, it was one of the few things to be found that does something similar to this, and thus the results are worth mentioning.

4) Research Questions and Hypothesis

The research questions for this study focus on the pedagogical implementation of SLA theories in these applications, and student perception of that implementation. There are five guiding research questions for this study:

1. Which of the following SLA theories being applied, in some way, by Duolingo and HelloTalk in their app/lesson design:
 - a. Level appropriateness and comprehensible input and output
 - b. Focus on form
 - c. Zone of proximal development and scaffolding
 - d. Interaction
 - e. Communicative language
 - f. Noticing
 - g. Motivation and gamification
2. What are the user perceptions of these applications' implementation of SLA theories and resulting instructional materials?
 - a. Strengths and weaknesses of the instructional materials/app design according to user perceptions.
3. What are users' primary goals and motivations for using these apps, or the reasons for maintaining or losing motivation?

The two applications are very different by nature. It is clear from a cursory glance at the application that Duolingo focuses on a rote, translation style of language learning, and probably utilizes some form of explicit feedback to facilitate learning while not focusing much on interaction. Meanwhile, a quick look at their website and app suggests that HelloTalk pivots in the other direction with a complete focus on interaction with peers. The interesting results will be how users feel about the materials and implementation of each application, and how these applications will fit in with learners' overall learning strategies. I hypothesize that, while the motivational and frustration issues mentioned in the literature review will cause many users to dislike the applications, overall sentiment towards both applications and their SLA implementation will be positive.

CHAPTER 4: METHODS

This study was qualitative in nature, with the purpose being to identify the implementation of SLA theory in these applications and ascertain the quality of that implementation. To achieve this, this study was split into three parts. The first part is a general review of the applications themselves and their features with the second and third parts consisting of a survey about student perceptions of the application's implementation of SLA principles, broken down into the different SLA concepts mentioned in the research questions, and followed-up with an interview for clarification and details. Before breaking down the steps of this study, it should be noted that this study will only be focusing on the *free* versions of the applications, as those are the versions that have the advantage of being readily available to everyone with a smartphone. While there are paid versions of both apps, those will not be investigated.

The decision to split this study into three parts was primarily to maintain validity. While it would have been possible and easy to simply *identify* the implementation of specific SLA theories in certain applications, making a value judgement on the quality of that implementation is not something that I can do without personal bias hindering that judgement. Thus, a general survey open to any and all users was decided upon to collect data on the perceptions of as many users as could be found to take it. Finally, the interviews were implemented to gain valuable details on the insights of some of these users. This multi-faceted approach provides the perspective of multiple users, and thus if there is a consistent pattern among the respondents it can be safely assumed that the value judgement is more valid than that of a single researcher.

The following segments will cover the details of each step of the research process. Part 1 will discuss the guidelines for the app review, and part two will discuss the details of how the

survey was created and organized. The third section will discuss the participant demographics and determiners for valid responses, and the last segment will discuss the guiding questions and organization of the interviews that were conducted.

1) App review and guiding questions.

The review of these applications is geared towards identifying the use of certain SLA principles in the features of the applications in question, rather than judging the quality of that implementation. This was done to determine what questions would be appropriate to ask survey participants about the applications. To maintain objectivity, the purpose of this analysis was determined to be only factual, with as little of a value judgement placed in the review as possible. Effectiveness of implementation is instead to be determined by user feedback.

Underlying all these lines of inquiry is the concept of communicative competence as defined by Hymes (1986), or in other words whether the instruction on forms helps students to utilize language in realistic situations with actual speakers. While it was determined that finding whether these apps help with communicative competence was important, a specific, singular feature that covers this concept would be difficult to create. However, it can be reasonable to expect that app design and lessons can revolve around realistic scenarios and useful language, and so this is a concept that will come up repeatedly when investigating the following concepts.

The first thing that was investigated was whether the applications had a way of determining level appropriateness. This line of thinking was guided by Vygotsky's Zone of Proximal Development (1978) and the idea that language is learned in stages of development, or "sequences of development" in the words of Lightbown and Spada (2017). In linguistic circles,

the concept of language appropriateness is better known as comprehensible input as argued by Krashen (2008). While the theory of comprehensible input is not perfect, whether the two applications in question contain a mechanism for determining linguistic ability, and whether both mechanisms are accurate and have a meaningful impact on users' learning experiences, may play a large role in how users feel about the applications' effectiveness. While the review will identify a mechanism for level determination, the survey questions will be used to determine whether students felt like they were placed at an appropriate level.

The second segment of the review delves into the type of feedback that the applications provide or can provide. Some concepts being focused on here are on grammar form and vocabulary learning. The basic question here being, first, whether the applications have a method or way of focusing on forms such as grammar, syntax, and pronunciation so that users are getting the form-based instruction they need as suggested by Hummel (2014). These questions and review segments are to be followed up by the survey and interviews to determine if the instruction on forms and vocabulary assisted the users with communicative competence, as determining whether the instruction helps requires the perspective of active or past users.

Also related to the review of feedback is whether these applications help the students notice their errors over time according to the theory of noticing, where there is a positive relationship between students noticing errors and overall learning (Mackey, 2006). While some of this can be investigated through a review of the applications, the best way to find out if users began noticing errors is through users' testimonials and feedback, and thus this will be primarily determined by related survey and interview questions.

A final feedback-related inquiry that is investigated via review is the way vocabulary is taught. There are a couple of pertinent questions in relation to vocabulary, first being related to

communicative competence, or whether the vocabulary taught is useful in real communicative situations. Second, and even more difficult to ascertain through the review, is whether the vocabulary is taught in a way that helps students remember it. This is possibly where the greatest potential for these apps comes into play, as the potential for language learning apps and vocabulary practice has been mentioned by multiple studies by Jones (2011) and Burston (2014) and has been confirmed by an earlier mentioned study by Walker (2015). These applications may show great promise in helping with vocabulary retention, and the usefulness of the vocabulary will be confirmed through the survey.

Next, a major point of investigation will be the aspect of motivation in relation to these applications. Gamification as described by Werbach (2014) is something that Duolingo is particularly known for (Rachels & Szapkiw, 2018). However, the motivation of students is complicated, and the reasons students stay, and leave, are likely to be just as complicated. However, it is still possible to find evidence and examples of gamification in a quick review and the efficacy of such tactics will be determined later.

Finally, options for social interaction offered by the applications will be explored in this review. While MALL is particularly known for limited, rote-learning programs and vocabulary learning (Burston, 2014), the guiding principle of HelloTalk is an app that might be capable of delivering on some of Burston's (2014) initially optimistic insights into the potential usefulness of MALL projects. While it will be easy to see whether social interaction is offered on these apps, the quality of the communities and the interactions therein can only be discovered through the survey data and interviews.

2) Survey Content

A survey with questions directly related to the topics in the review was administered to participants to collect data on user perspectives of the features and implementations of SLA theory in these apps. Each SLA principle in the SLA concepts mentioned in the research questions has questions dedicated to eking out relevant user perspectives regarding them, involving at least 2 questions for each research question SLA category. Many of these questions are direct follow-up questions to the app review. The following segments will describe the lines of inquiry and the related questions in the survey.

The survey was distributed through Qualtrics and sent to participants using social media, Reddit, and emails. Some participants were recruited through word-of-mouth. Some challenges in the distribution of this survey were the low initial responses and a slew of incomplete responses as well as invalid responses. As mentioned before in the Participants section, there were 200 responses but only 101 were considered valid based on the checklist provided in that section. Many of these participants either did not answer any questions at all, or only answered the demographics questions without giving any good feedback on the applications themselves. An additional struggle was the unpredictability of getting forum users to respond to the survey. Many times, an announcement would be placed on a forum for no respondents to show up on the survey, and other times a survey post would be followed by dozens of responses. Some of the most fruitful forums for response numbers were r/languagelearning and r/englishlearning on Reddit.

The survey was separated into three sections for the participants. The first part was permissions and demographic information, while Duolingo and HelloTalk were two separate sections of the survey that only showed if students indicated that they had used the app. The

survey contained sixty-four questions total, but only respondents who used both applications would have answered all the questions.

Demographics

The first questions in the survey aimed at learning demographics details of participants. The first questions of these asked about age and education levels to see the distribution of participants in those aspects. Second, the language that participants spoke at home as a child and the language that they learned using the applications was asked as well, to determine the language demographics of the users that responded.

Frequency and Duration of Use

Before getting into the SLA theories themselves, some questions were determined important for general knowledge of the user base. It is important to note how long most survey takers had been using the applications, and how often it was used. To account for this, each app section of the survey starts off with a series of questions to determine for how long the users used the application and how frequently they used these applications. The frequency and duration of use ties into the concept of attrition explored by Tuncay (2020) and can give additional insight into the overall sentiment towards an application outside of the effectiveness questions later on in the survey.

For how long did you use Duolingo? (Or have you been using)

A few days

1-2 weeks

Several weeks

2-6 months

6 months to a year

More than a year

Figure 1: Duration of use question

How often did you, or do you, use Duolingo?

Daily

4-6 times a week

2-3 times a week

Once a week

Rarely

Figure 2: Frequency of use question

The questions for frequency and duration of use shown in figures 1 and 2 are kept simple so that Users can give general information. The first question (figure 1) asks how long the user used the application, and the second question (figure 2) how often the user used the application when they were using it. It is hoped that attrition rate of the applications can be extrapolated from the information in these two questions, and coupled with the attrition rate, the questions later in the survey can help to determine the reasons for those rates.

Macro Questions

A second set of questions that were asked that were only loosely related to the SLA theories explored in the questions below were a set of questions gauging what students considered to be important when using these applications, as well as follow up questions gauging how helpful or effective the application was at helping the learner with those skills. The categories for both of these questions included the following:

- Improving writing

- Improving reading
- Improving listening
- Improving translation skills
- Improving grammar skills
- Improving vocabulary
- Supplementing in-person classes
- Supplementing conversation practice with a language exchange or tutor.

For the personal importance of each category, participants rated each category from extremely important to not important at all on a five-point scale. For effectiveness in these categories, participants then rated each category from extremely effective to not effective at all on a similar five-point scale.

These questions were designed to gauge the overall utility of the applications for the users. It is unlikely that all of these users are using only these applications for their language learning, and thus it is important to gauge what students intend to get from these applications when going into them. Additionally, the two applications are extremely different in nature, and the reason for using the application, and perceived effectiveness of the application for aspects that are important to the learner may influence motivation as well, which will also be explored in more detail later in the survey.

Level Appropriateness & Comprehensible Input and Output

The next section begins with questions directly related to the SLA theories being discussed. There were a few different chunks of questions here. Duolingo has a level test available for users who may have previous experience with a language, so the first chunk of questions is dedicated to those learners and is only available to participants who answered the survey for Duolingo. If participants for Duolingo indicated that they took Duolingo's placement

Did you try to skip ahead of the level that Duolingo recommended after taking the placement test?

Yes

No

Did you decide to study at a lower level than the placement test recommended?

yes

No

If the answer to either of the previous questions was yes, can you explain why you decided to skip ahead or start at a lower level?

test, they were prompted with the series of questions shown in figure 3. These questions were designed to get additional details from the participants about what they may have done as a response to the level test. The final short answer question was intended to give students a chance to

Figure 3: Follow-up questions regarding the Duolingo level test

explain the previous two answers, as

students indicating that they went back or skipped ahead does not specifically mean that they

disagreed with the level test assessment. HelloTalk did not include this line of questioning, as the program does not have a level test, and instead prompts learners to self-report their levels.

The next series of questions in this category were made to get details on the level appropriateness of specific input and output. These two questions asked students how they felt about the input and output broken down in the following categories:

- Grammar
- Vocabulary
- Written Language
- Spoken Language/Audio
- Overall Input/Output (with layman explanations of the ideas in parenthesis)

These categories were broken down into seven ratings ranging from extremely appropriate, explained as not too difficult or too easy, to extremely inappropriate, meaning either so difficult that it was impossible or so easy that the user did not learn anything, as explained to the survey takers. Both survey sections for Duolingo and HelloTalk included this line of questioning. However, the questions for HelloTalk specifically ask about the level appropriateness of the language given to learners by their *language partners* specifically, as the lessons that may be available to some students are not a part of the free versions that are being researched in this study.

Focus on Form (FoF)

The FoF set of questions involved a series value judgement questions regarding lessons and explanation of grammar, vocabulary, and pronunciation and related short-answer questions

for participants to provide details. The first question in the series was a basic value question asking participants how well, on a five-point scale from extremely well to not well at all, the application helped the students learn grammar forms, which was then followed up by short-answer questions that gave students a chance to explain, which was similarly done with vocabulary as well. Duolingo has some grammar instruction in its app shown later in this paper in the analysis of the applications, while the questions for HelloTalk specifically focus on the grammar discussions with the language exchanges in the application.

ZPD and Scaffolding

This line of questioning in the survey is entirely dedicated towards Duolingo, as HelloTalk lacks guided instructional material on the free version of the app and therefore would not have the lesson planning needed to implement scaffolding. This chunk involved two primary questions, with the first question being how well users felt the materials covered in each lesson prepared participants for upcoming lessons with the range being from extremely well to extremely poorly on a five-point scale to determine scaffolding, and with the second question asking students if they felt new lessons adequately reviewed previously covered materials to reinforce learning, because forgetting previous material is a normal part of the language learning process (Lantolf & Alijaafreh, 1995). Both questions were designed to determine if lessons built upon each other and thus if scaffolding and the concept of the ZPD was present in the program. Coupled with questions on level appropriateness, inferences can be made on whether students are working on the target language within their ZPD. There were no follow-up short answer questions in the survey to cover this segment but scaffolding and level progression was brought up in a lot of the interviews to accommodate for this.

Interaction

For Duolingo, there were no survey questions that referred to interaction, because the program does not attempt to provide a similar interactive language learning environment like HelloTalk as a platform. For HelloTalk however, the principle of the application is entirely through interaction, and all the subsequent questions about other aspects of the app indirectly revolve around this design. As a result, specifically asking students how their interactions with their peers went into the questions related to all of these theories. So, while there are no questions specifically for interaction for HelloTalk, all of the data for the HelloTalk survey can be looked at in the lens of interaction as well.

Communicative Language

Questions about communicative learning were distributed throughout the survey in questions about grammar and vocabulary. Most of these questions expanded upon the effectiveness questions by asking if participants felt that the language being taught was useful for certain contexts situations. The first related question for both apps was about how students felt about how the app helped with grammar in specific contexts: casual conversation, formal conversation, speeches, writing emails and notes, and writing formal essays and articles on a five-point scale from extremely well to not well at all. The second related questions were about vocabulary effectiveness in scenarios on a scale of five ratings from extremely effective to extremely ineffective, and the range of vocabulary that students felt they learned on a scale of five points from extremely broad to extremely narrow,.

Noticing

The questions on noticing included a series of three questions for each app; one that was an overall value of how students felt using the app helped them notice their errors in grammar, vocabulary, and pronunciation, and the other two were questions about how students felt about the app or their language exchanges specifically showing mistakes and a short answer question to

How helpful do you feel Duolingo was for helping you to notice your own mistakes in the following categories?

	Extremely helpful	Somewhat helpful	Not helpful at all
Grammar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vocabulary	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pronunciation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

How effective do you feel Duolingo was at specifically showing you your mistakes when you got a question or exercise wrong?

Extremely effective

Very effective

Moderately effective

Slightly effective

Not effective at all

What did you like or dislike about how Duolingo showed you your mistakes or corrected you?

Figure 4: Noticing-related questions in order as they appear in the survey.

follow up. Figure 4 shows how these questions appeared in order. These questions were also able to give valuable insight into focus on form, but most importantly for this segment these questions were designed to determine if learners were able to begin identifying their own mistakes in the app. During the interviews, additional questions were asked about whether this application helped participants notice their own mistakes outside of the program as well as follow up questions about whether Duolingo's style of correction was useful, and if language exchanges in HelloTalk were helpful in this aspect.

Motivation and Gamification

Questions about motivation were the same for both applications. Questions about motivation came in a series of four questions, starting with a general question about extrinsic motivation for the students shown in figure 5. This question about extrinsic motivation was to

	Extremely important	Very important	Moderately important	Slightly important	Not at all important
Work	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
School	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Personal interest in English or languages in general	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Communicating with friends or family	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Being able to participate in online interaction or gaming	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Travel	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 5: Question regarding what students consider important when using the app to learn the target language.

gauge what brought learners to the application to begin with. While gamification might turn out to be a big factor for why users continue to use Duolingo, chances are some form of extrinsic motivation encouraged many learners to start the app to begin with. After gauging extrinsic motivation, the following three questions attempt to gauge how long users stayed motivated using the application and then a final open-ended question asking what kept them using the app if they stuck with it.

3) Survey Participants

The survey was advertised online and administered through Qualtrics for a period of roughly three months. Participants for this survey were recruited almost entirely online, through forums and social media communities with the occasional recruitment through word of mouth. Participants in this study were completely anonymous when taking the survey and were not required to participate in interviews after the survey was taken. Those who did wish to participate in interviews were only required to provide an email for communication, and attended the interviews through Zoom, giving many more options for anonymity if they wished to not give their name or any other identifying information. and while well over 200 responses were recorded overall, only 101 were considered valid when data was collected. Valid responses were determined through a quick review of all responses that participants left. The reasons for a response to be disqualified included the following:

1. The participant did not respond to any of the qualitative questions regarding the applications. I.E., *only* answered the demographics questions and nothing else.
2. The participant did not indicate their first language and the languages learned using the applications in question.

3. The participant indicated that they have no experience with the applications being investigated.
4. Multiple responses from the same IP address were deleted, with the most complete response being kept.
5. Partially completed responses were kept if they provided at least *some* valuable responses to the qualitative questions regarding the applications, with the absolute minimum questions required being a response to their personal motivations for using the application, and how effective they felt the application was in helping them with their reasons and goals for learning.

Some basic demographic questions were asked about the students. For gender, most participants, fifty-eight, were female, while thirty-seven were male. Four participants were non-binary and two preferred not to answer. Age was also asked, and the most common age groups for the study were between twenty-five and thirty-four, with the second most common group being between eighteen and twenty-four, 39.60% and 32.67% of participants, respectively. No one above 65 took the survey. Finally, the education level was asked, which was by far the most varied demographic of the study. While many users had a four-year degree, thirty-eight users, the other demographics are split fairly randomly, with professional degrees being the second largest group at 21 participants. Figure 6 shows the education split among users who took the survey. While university educated users make up the bulk of the participants, there was still a significant portion of the survey takers that had some college or less.

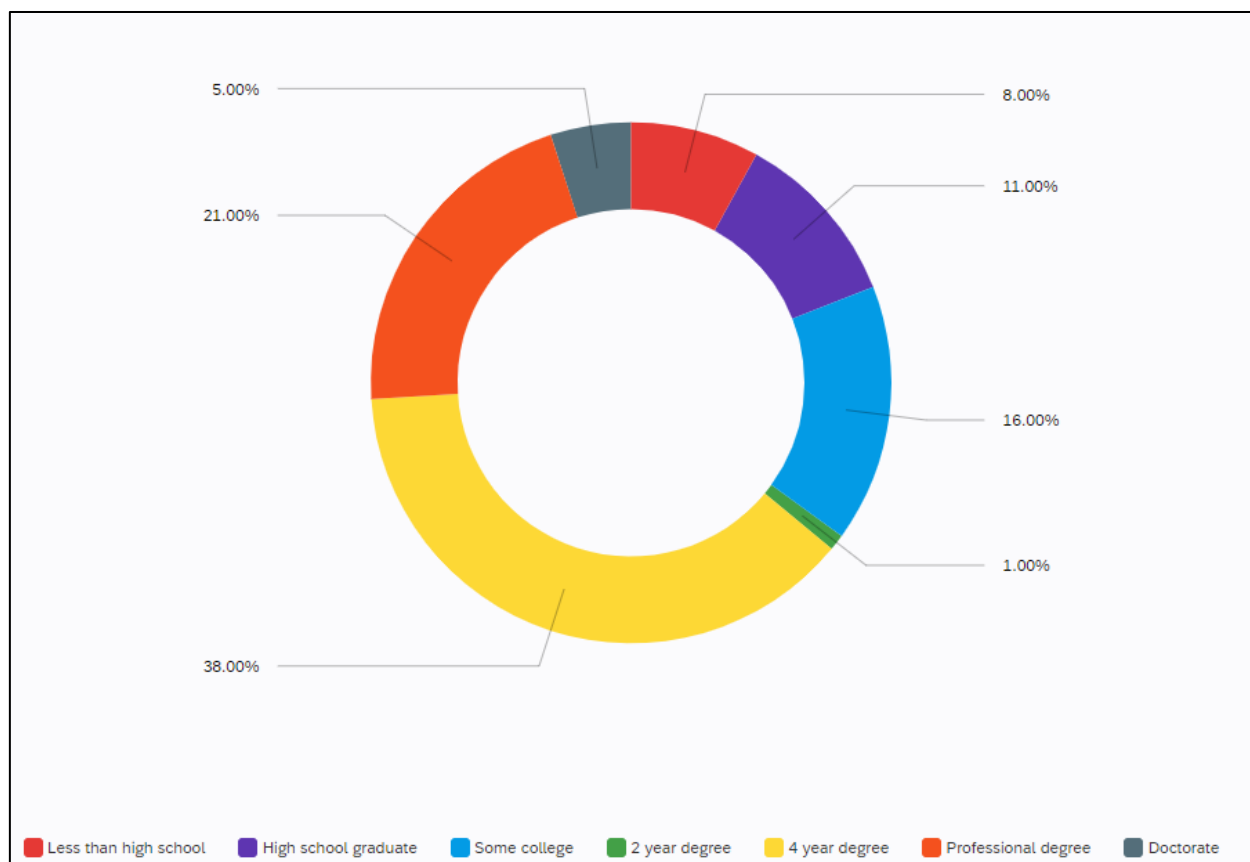


Figure 6: Participants divided by education level.

Outside of demographics, participants were asked to self-determine their proficiency levels in the target language when using these apps. The levels of proficiency in the survey were based on the CEFR levels in the European Framework and contained detailed descriptors next to them in the survey to help students determine their level as shown in the figure. Figure 7 shows a

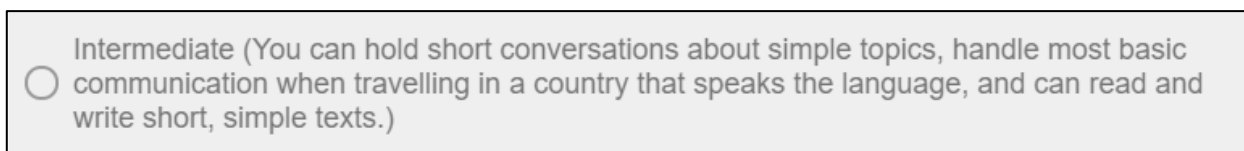


Figure 7: Example question for participants' self-reported ability in target language.

general level that students can choose from with a detailed description next to it. This level is based on the CEFR B1 level, and the description is to give participants a specific idea of what the ability that this level should entail. Additionally, students were asked what the language

spoken at home when growing up was, and what language was learned using the applications being investigated.

The CEFR was chosen as the guidelines of these level determinations because of the solid foundation that the CEFR has laid in determining language levels in the European Union. The US lacks a solid consensus on level determination while the Council of Europe's level determiners are very specific and relate to detailed skills and stages of ability to use a language. These descriptors range from the lowest level, A1, which is described as

Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help. (Council of Europe)

To the highest level, C2, which is described as

Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments, and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations. (Council of Europe)

All the levels in-between contain similar, detailed descriptions that give a clear idea of what each level should be capable of. The six levels of the CEFR are split between three categories: beginner, intermediate, and proficient, with both high and low segments of each section. For example, A1 would be a low beginner while A2 would be a high beginner. In the same vein, B1

would be low intermediate and B2 would be high intermediate. Because of this clarity and the detailed descriptions, the descriptors for the different CEFR levels were simplified for the average survey taker and implemented into the level determination question for more accurate responses. While self-reporting for something like language levels can never be 100% reliable, it is hoped that, with accurate descriptions, survey takers would be less susceptible to exaggeration or downplaying of personal ability and try to be as accurate as possible. At the absolute minimum, solid descriptions such as these avoid students incorrectly labeling their own linguistic ability due to confusion.

A majority of participants rated themselves as some sort of beginner, 45 in total, in the target language that they were learning using the applications, either absolute beginner, 19 participants, or high beginner, 26 participants. The second largest group of participants rated themselves as some form of intermediate, 36 in total, with most of the respondents in this group reporting themselves as lower-intermediate, 24 participants, rather than upper intermediate, 12 participants. Comparatively few participants considered themselves proficient in the target language, 20, with only 6 of those participants claiming the highest level of proficiency, or a C2 level.

4) Focus Groups and Interviews

Participants for interviews were recruited from survey participants who indicated that they were willing to take part in an interview and left an email for contact. While the original plan was for focus groups that could interact with users of similar backgrounds, such as English first-language (L1) speakers who used Duolingo, or Spanish L1 speakers who used HelloTalk and Duolingo, with categorization meant to focus on the apps that the respondents discussed in their surveys, the largest group by far was English L1 users who used both HelloTalk and

Duolingo. Additionally, English users of the applications were the only participants to answer the call for interviews when the email was sent out, with the exception of one participant who was a Portuguese L1 who used Duolingo to learn Spanish. In another severe blow to the original plan, the disparate time zones and busy schedules of the participants made it difficult to schedule anything other than individual interviews, so only interviews were conducted rather than focus groups.

The purpose of the interviews was to give participants a chance to elaborate on the comments and responses to questions in the survey. Ideally, students were to do this in groups so they could also freely discuss their responses together and compare experiences. In the end, the interviews were done entirely one-on-one. Some questions were specific to the participants, as participants' responses were identifiable through their email, and as such specific details or clarifications could be asked of these users. However, there were also a series of guiding questions that were geared towards keeping the conversation focused if there were no participant-specific lines of inquiry. These guiding questions for the focus groups and interviews were:

1. How often was the material in the app too easy or too difficult for you? What about the materials wound up being difficult?
2. Were you able to understand most of what you heard and read when using the app?
3. Were the writing/speaking exercises within your skill level? Were they ever too difficult or too easy?
4. Can you think of a situation where the language you learned using this application helped you in a real-life situation?

5. In general, was this language useful for real-life situations? If so, what situations did it help with the most?
6. At any point were you able to notice your errors before being corrected in the application? Were you eventually able to notice errors that you made in the language outside of the app?
7. What motivated you to keep using the app? Was it mostly personal reasons or did the app provide some sort of motivation for you?
8. (For HelloTalk only) Were the social interactions on the application helpful to you in developing your language? Why?
9. (HelloTalk only) Did you feel that your language exchanges gave you good feedback?
10. (HelloTalk only) Did anything frustrate you about the language exchanges that you talked with?

The conversations could, and did, frequently deviate from these specific questions depending on the feedback of the user and their individual experiences with the apps, with follow-up questions for more detail and clarification frequently being asked.

In general, an hour was budgeted for each focus group. However, the time for each was entirely dependent on how much the participants had to say about the subject. The vast majority of interviews lasted around 30 minutes and were recorded on Zoom with transcripts for the sake of note taking. Participants were informed that their video and audio would not be used as evidence in the study, and that any quotes used would utilize alias names rather than their real names. Aliases were determined using a random alias generator online or by participant request.

CHAPTER 5: ANALYSIS & DISCUSSION

In this chapter, each section will pertain to one of the SLA principles asked about in the research questions. There are nine sections to this analysis, with seven of them directly pertaining to the research questions and two parts asking relevant background information about frequency and duration of use as well as what learners consider to be important when using these applications and perceived overall effectiveness. These first two parts are intended to give some extra context that might not be directly related to the SLA concepts being researched but might influence overall responses to the survey.

Each segment will include an overall review of the relevant app features for the SLA concept being discussed. Some SLA theories are more relevant to specific apps, which will be broken down in the dedicated sections. Additionally, each segment will go over response data from the survey for questions relevant to the segment's topic. Finally, each segment will end in a general discussion about the data and will incorporate short answers from participants as well as quotes from the interviews to provide context to the data and assist with the overall analysis.

Before breaking down the responses of the survey, it should be noted that, while 101 survey takers responded, not all survey takers completed the survey. Some participants left valuable feedback in the opening sections of the survey, answering macro-level questions about overall learning goals and whether the application in question helped with those goals, but failed to answer some of the more in-depth questions. Additionally, there were wide discrepancies between the number of people responding to each application. Eighty-eight people participated in the Duolingo survey, with only fifty-two of those completing the survey in its entirety to the last question. Meanwhile, only sixteen people participated in the HelloTalk portion of the survey, with only twelve completing it in its entirety. While it is disheartening that so few respondents

completed the HelloTalk survey, a few respondents who offered to do interviews had experience with both applications, giving a good chance for comparison in the interview portion of this study. Additionally, HelloTalk respondents were much more responsive on the short answer questions in the Survey, giving a good variety of perspectives to back up with responses to the survey questions.

Another thing to note about the survey is that some of the SLA principles that are being investigated in this research are more applicable to one application over the other and vice versa. For example, Duolingo does not have many chances for interaction built into the program, but HelloTalk is built completely around that aspect of SLA. Additionally, HelloTalk is made primarily for informal language learning through language exchanges, so the questions about HelloTalk, while covering the same general concepts, will be relating most of these principles to the language exchanges that users interacted with on the application. In short, while the survey questions will ask users to evaluate the instructional materials of Duolingo, the same questions for HelloTalk will be more about whether language exchanges on HelloTalk were able to help in the same way. HelloTalk provides the tools to help with all of these principles, but the survey is more about learning whether or not the language exchanges utilize them and if users find the subsequent interactions useful.

1) Frequency & Duration of Use.

The results of the frequency and duration of use questions for Duolingo and HelloTalk vary for the two applications. As can be seen in figure 8, a majority of participants, 29.55%, used Duolingo for more than a year while the second largest portion of survey takers, 23.86%

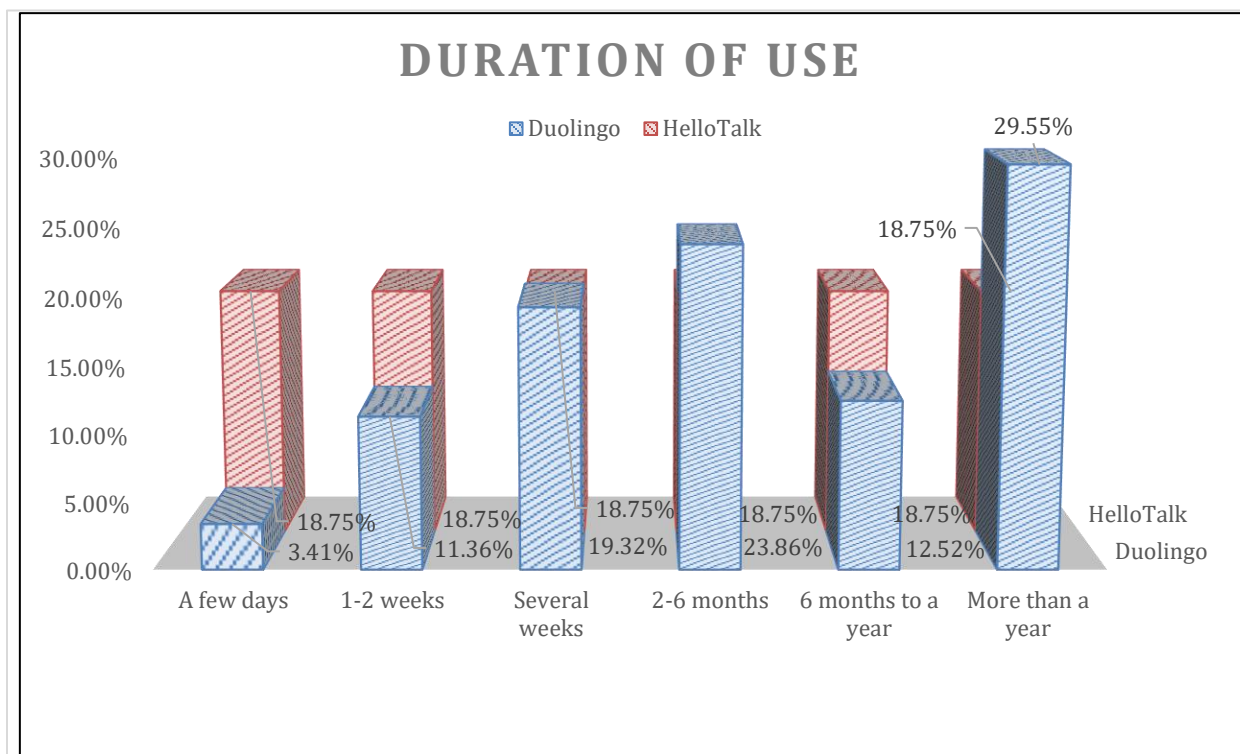


Figure 8: Comparison between responses to Duolingo & HelloTalk duration of use question by percentage of participants.

participants, had used the application for between two to six months. Around 12.50% of participants had been using the application from 6 months to a year, 19.32% of participants for several weeks, and 11.36% of participants for between one to two weeks. 3.41% of participants had only used the application for a few days to a week. In comparison with Duolingo respondents, the split between length of use for HelloTalk among users is pretty even, with 25% of respondents stating that they only used it for 1-2 weeks, with all other categories such as several weeks, 2 to 6 months, 6 months to a year, and more than a year having 18.75% of respondents each.

For daily use, Figure 9 shows another disparity between the two applications, showing that users were more likely to use Duolingo daily. 42.05% of them used Duolingo daily, with 20.45% of participants saying they used it between four to six times a week, and 15.91% of

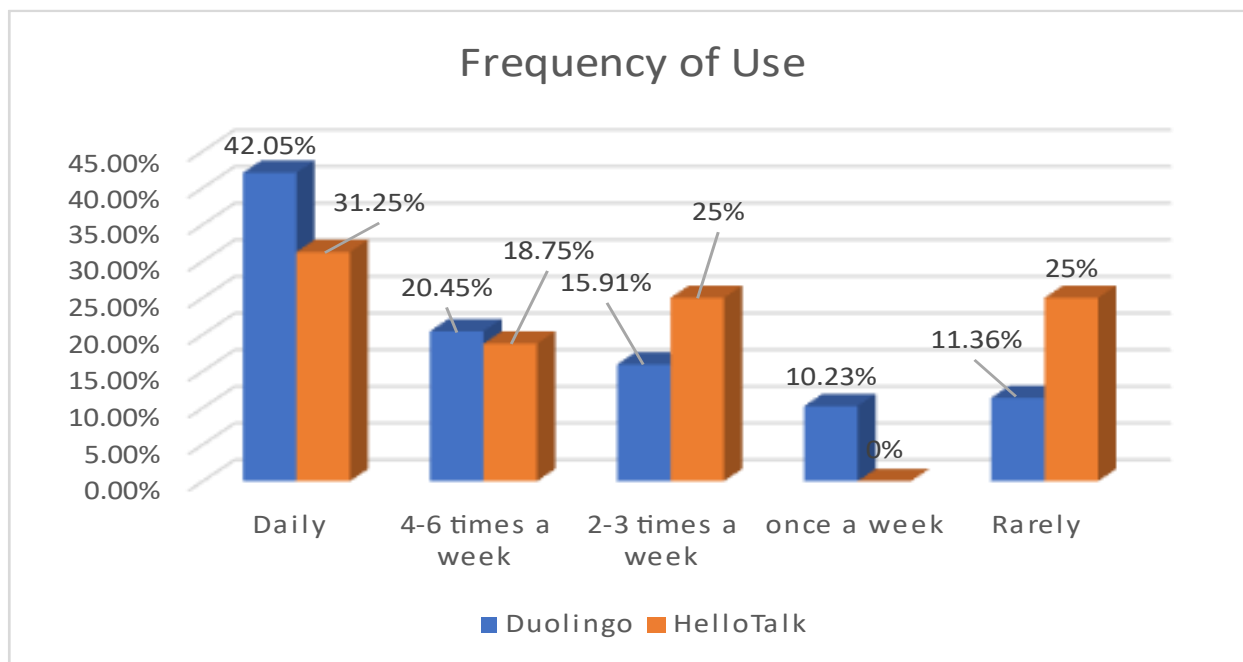


Figure 9: Comparison between frequency of use between Duolingo and HelloTalk.

participants reporting that they used the application two to three times per week. 10.23% of participants stated that they utilized the application once a week, and 11.36% of participants stated that they used it rarely. Of the respondents who used HelloTalk, 31.25% used it daily, 18.75% used the application four to six times per week, and 25% used it rarely.

From this data it is clear that Duolingo was more likely to be used more frequently, and especially for a longer duration, by its users. The percentage disparity in both aspects in favor of Duolingo suggests that something may have pushed users away from HelloTalk more often than not.

Additional survey data showed that a larger percentage of participants no longer use HelloTalk to this day. While 59.09% of participants stated they no longer use Duolingo,

HelloTalk fared significantly worse in attrition rates, with 81.25% of participants who used it previously stating that they no longer use it. Later in this analysis, in the motivation section, further survey data and testimonials from short-answer questions and interviews will shed some light on the reasons behind this.

2) Duolingo Macro Questions: What's Important For Learners and Overall

Effectiveness.

For Duolingo, what mattered to students was varied. For what learners found important to their learning, a few categories stuck out, which can be seen in figure 10. First, improving

Categories	Extremely important	Very important	Moderately important	Slightly important	Not at all important
Improve writing	6.82%	17.05%	30.68%	22.73%	22.73%
Improve reading	19.54%	29.89%	28.74%	12.64%	9.20%
Improve listening	18.39%	29.89%	28.74%	14.94%	8.05%
improve translation skills	17.24%	26.44%	22.99%	22.99%	10.34%
improve grammar skills	19.77%	25.58%	26.74%	15.12%	12.79%
improve vocabulary	37.93%	41.38%	13.79%	5.75%	1.15%
Supplementing in-person classes	6.90%	11.49%	18.39%	13.79%	49.43%
Supplementing conversation practice with language exchange or tutor.	6.90%	10.34%	11.49%	19.54%	51.72%

Figure 10: What users considered important when using Duolingo.

vocabulary was the most common goal that students labeled as extremely important or very important, with 37.93% and 41.38% of respondents stating so respectively. Improving reading and listening skills were the next two top categories that were frequently listed as very important or extremely important, being tied for very important at 29.89% of respondents each, and being extremely close in the extremely important category, with 19.54% of respondents saying that reading was extremely important and 18.39% saying listening was extremely important.

Grammar was also consistently labeled by most respondents as being at least moderately important or more, with 19.77% of respondents declaring it to be extremely important, 25.58%

of respondents saying that it is very important, and 26.74% saying that it was at least moderately important. Two categories ranked low in the scale of importance for Duolingo users, with supplementing in-person classes being declared as not at all important by 49.43% of respondents and supplementing conversation practice with a language exchange or tutor being stated as not at all important by 51.72% of respondents. While some students declared these aspects as being extremely important and very important, those ratings were not significantly high for either category.

Comparatively, what users found important when using HelloTalk is different than Duolingo at first glance. As can be seen in figure 11, a higher percentage of participants, 25% indicated that conversation with a language exchange or tutor was a major reason for using

Categories	Extremely important	Very important	Moderately important	Slightly important	Not at all important
Improve writing	18.75%	31.25%	18.75%	12.50%	18.75%
Improve reading	12.50%	43.75%	25.00%	0.00%	18.75%
Improve listening	12.50%	12.50%	50.00%	6.25%	18.75%
improve translation skills	12.50%	18.75%	25.00%	6.25%	37.50%
improve grammar skills	18.75%	25.00%	31.25%	18.75%	6.25%
improve vocabulary	25.00%	37.50%	25.00%	6.25%	6.25%
Supplementing in-person classes	0.00%	26.67%	20.00%	20.00%	33.33%
Supplementing conversation practice with language exchange or tutor.	25.00%	12.50%	25.00%	18.75%	18.75%

Figure 11: What students considered important when using HelloTalk.

the application. Many participants indicated a desire to improve reading and writing as well as grammar skills, but at different levels of importance compared with participants in the Duolingo survey, with over half of participants putting those three categories as moderately important or higher. There were no aspects of learning that were consistently low on the list, but improving translation came out as the least important in general, with 37.50% of the participants regarding

it as not at all important. Supplementing in-person classes was also unpopular, with 33.3% of the students considering it not at all important.

While the reasons for using the applications are spread out for respondents, the responses for how effective learners found Duolingo overall in the mentioned categories tend to be in the middle. As can be seen in figure 12, almost all categories for Duolingo have a majority of

Category	Extremely effective	Very effective	Moderately effective	Slightly effective	Not effective at all
Improve writing	3.57%	8.33%	30.95%	35.71%	21.43%
Improve reading	9.76%	13.41%	36.59%	31.71%	8.54%
Improve listening	10.84%	14.46%	30.12%	25.30%	19.28%
improve translation skills	6.10%	19.51%	36.59%	29.27%	8.54%
improve grammar skills	6.10%	8.54%	39.02%	28.05%	18.29%
improve vocabulary	15.66%	31.33%	33.73%	16.87%	2.41%
Supplementing in-person classes	4.88%	4.88%	18.29%	17.07%	54.88%
Supplementing conversation practice with language exchange or tutor.	2.47%	2.47%	19.75%	17.28%	58.02%

Figure 12: How effective Duolingo users considered the app to be in categories related to importance.

responses as moderately effective, with a few exceptions; Improving vocabulary is almost tied between very effective and moderately effective, with 31.33% of respondents stating it was very effective and 33.73% of respondents stating that it was moderately effective. Supplementing in-person classes and supplementing conversation practice with a language exchange or tutor are both mostly stated as being not effective at all, with over 50% of participants stating so for each.

Another difference between Duolingo and HelloTalk is that HelloTalk was rated consistently low in the categories charted in figure 13. Overall, in terms of how useful users found it for the aspects of learning. Some users, 33%, found it very effective for improving translation skills, and 25% found it very effective for improving reading, grammar, and for

Category	Extremely effective	Very effective	Moderately effective	Slightly effective	Not effective at all
Improve writing	0.00%	8.33%	33.33%	41.67%	16.67%
Improve reading	0.00%	25.00%	33.33%	16.67%	25.00%
Improve listening	0.00%	8.33%	25.00%	16.67%	50.00%
improve translation skills	0.00%	33.33%	0.00%	25.00%	41.67%
improve grammar skills	0.00%	25.00%	25.00%	16.67%	33.33%
improve vocabulary	8.33%	16.67%	16.67%	33.33%	25.00%
Supplementing in-person classes	0.00%	0.00%	36.36%	9.09%	54.55%
Supplementing conversation practice with language exchange or tutor.	8.33%	25.00%	16.67%	8.33%	41.67%

Figure 13: How effective HelloTalk users considered the app to be in categories related to important.

language exchanges, but even those skills were not consistently high. Even for the aspect of learning that was rated comparatively high on the list of priorities for students to Duolingo respondents, the ratings were fairly low in terms of overall effectiveness; conversation practice with a language exchange or tutor was rated as not effective at all by 41.67% of respondents. Helping to improve translation skills was also poorly rated with 41.67% of respondents saying that it was not effective at all. Other than improving writing and vocabulary, all of the aspects have their highest response rate in the not effective at all rating.

The differences in what was important for users when using the apps is not all that surprising, considering that the two applications are very different in nature. However, what is surprising is that the users of HelloTalk considered the app and their language exchanges on it to be consistently less helpful in helping with the categories they listed as important to them. This lack of perceived effectiveness might be an additional explanation for the high attrition rate that

affects mobile apps in general Tuncay (2020), and the particularly high attrition rate of HelloTalk in conjunction with the low ratings of effectiveness gives this idea some credence.

Later in this analysis, the reasons for the low perception of effectiveness will be discussed as we delve into the SLA theories and how effective students found the implementation to be. Related issues and user perspectives on HelloTalk and its community will come up in the motivation section of this study. Although comparatively better in terms of attrition rate, Duolingo still has a high attrition rate, with almost 60% of respondents reporting to not use the application anymore, and most judgements of effectiveness trending towards the middle, there are likely elements of SLA implementation on Duolingo's application that are driving users away as well, and maybe some aspects of it that are keeping users in the app, which will be discussed further in the focus on form and motivation sections.

3) Level Appropriateness & Comprehensible Input and Output

Duolingo starts off its application experience with a question that seeks to identify the learner's level or previous experience. If the user chooses the second option shown in figure 14, claiming to already know some of the language being learned, then they will be prompted to take a level test that the application uses to determine what overall level and unit the learner will start in. The primary exercises for this test are translation exercises where, on the phone version being focused on in this study, users choose words from a word bank to form sentences that translate the given sentences from their first language to the target language and vice versa. While the

phone version utilizes these word banks, the computer edition allows learners to choose between word banks and typing their answers manually.

This test will give the learner a mix of simple and complex sentences utilizing different vocabulary and grammatical structures, presumably determining level through an analysis of which structures and vocabulary the test taker got wrong. However, what determines the level is not entirely clear; the review of this application included completion of both the Spanish and English tests for comparison, getting only one question wrong in each, but ending with slightly different results. When taking the Spanish test, a question closer to the end of the test, where the structures seemed to get more complex, was entered incorrectly. For the English test from Spanish, a question closer to the beginning of the test was answered incorrectly for comparison. The Duolingo placement test determined that the level of the English test should start in much later in unit two, while the Spanish test should

start much earlier in unit two. Meanwhile, perfect scores on both tests resulted in similar results, with both the Spanish and English tests being well near the end of unit two. Additionally, As would be expected, getting all questions wrong on the test resulted in starting from zero. A test-taker who fails the first set of questions outright in both cases will be taken out of the test and



Figure 14: Given options when starting Duolingo for the first time.

started at the beginning, suggesting that the test may start out with structures determined to be simpler and advancing to more complex content as the test taker successfully completes question sets.

Of the learners who responded to the survey, 56.63% stated that they utilized Duolingo's placement test and commented on the following questions. A significant majority of these

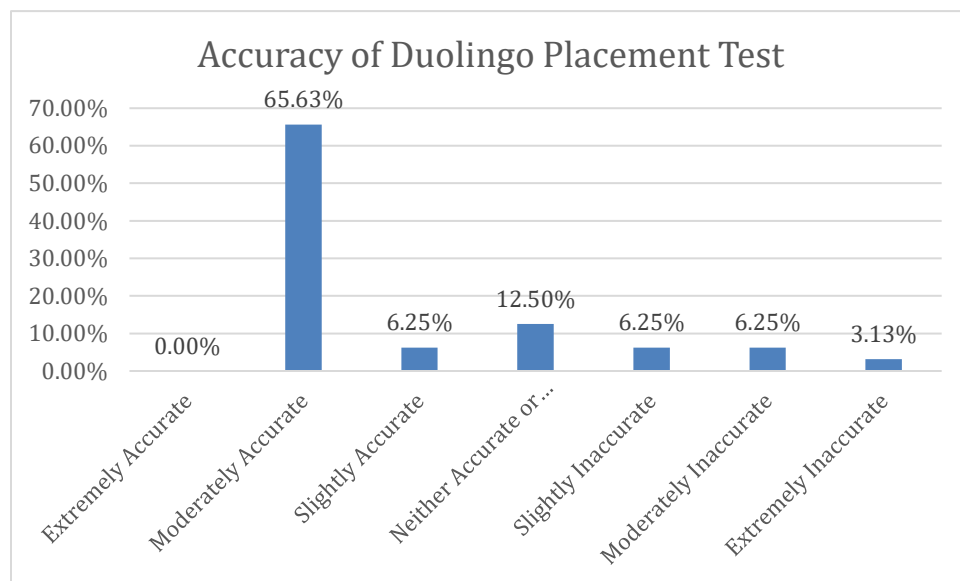


Figure 15: Accuracy of Duolingo placement test according to users.

respondents, 65.63% of respondents, stated that the placement test was moderately accurate, while none stated that it was extremely accurate, and the remaining

responses are relatively evenly distributed through the value judgements (Figure 15). 25% of learners attempted to skip ahead of the test by utilizing the quizzes available for each level, and 25% of learners decided to start at a lower level than the test assigned them.

The reasons for students skipping ahead or starting at a lower level are varied. Two students who answered a related short-answer question stated that they enjoyed reviewing previous material, while one student stated that “. . . there were things in the course that [they] didn't know, but [was] still placed ahead of.” as their reasons for going back through the material. Some other users decided to skip ahead; one survey taker commented that “[they] skipped ahead because otherwise [they] would have had to complete a lot of levels of things that

[they] already knew.”, which was echoed by two other similar responses to the short-answer question. Overall, most responses indicated that users were okay with their placements in the program, but either went back to review what they knew or skipped ahead for a challenge, depending on personal preferences. These users indicated that they were somewhat satisfied with the level that the placement test put them in, suggesting that the procedures for the placement test are fairly accurate with some outliers.

Overall, for Duolingo, students were satisfied with the level appropriateness of the input they received from the application regardless of level placement. Figure 16 shows that users

	Extremely appropriate	Moderately	Slightly	Neither	Slightly	Moderately	Extremely inappropriate
Grammar	26.92%	28.85%	17.31%	11.54%	7.69%	5.77%	1.92%
Vocabulary	30.77%	28.85%	21.15%	7.69%	3.85%	5.77%	1.92%
Written Language	23.08%	40.38%	13.46%	7.69%	5.77%	5.77%	3.85%
Spoken language/ audio	44.23%	17.31%	15.38%	7.69%	5.77%	3.85%	5.77%
Overall input (Examples of language, not what you created or produced.)	29.41%	25.49%	15.69%	19.61%	1.96%	1.96%	5.88%

Figure 16: Level appropriateness of input in Duolingo lessons according to users from Extremely appropriate (left) to extremely inappropriate (right).

considered the input they received from the application to be moderately appropriate to extremely appropriate. Spoken language was the strongest of the categories, with 44.23% of the participants stating that the audio input was extremely appropriate. The two other specific categories, grammar and vocabulary, were also rated consistently high by participants, with 26.92% and 30.77% rating them as extremely appropriate respectively, and 28.85% rating both of them as moderately appropriate. Written language was also rated consistently high, with 23.08% of students rating the overall written language as extremely appropriate and 40.38% rating the overall written language as moderately appropriate. Overall, the input was rated by 29.41% of learners as extremely appropriate and 25.49% of learners as moderately appropriate. For the appropriateness of output, participants for Duolingo rated it similarly high.

HelloTalk's method of determining linguistic ability generally relies on users self-reporting their levels. When choosing a second language the user will be prompted with five levels, which is beginner, elementary, intermediate, advanced, and excellent, with no descriptors to help users with making their determination. It is interesting to note that HelloTalk has instructional materials behind a paywall that uses the CEFR scale from A1 to C2 as can be seen in figure 17. However, the free version of the app does not utilize this same scale when determining personal ability.

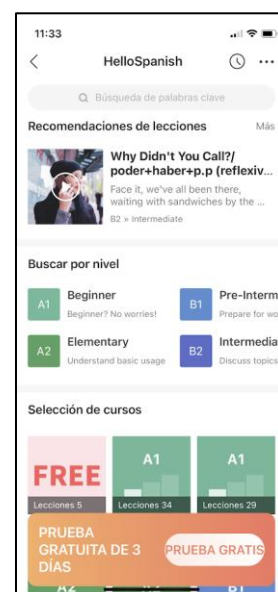


Figure 17: HelloTalk lessons and educational materials behind a paywall.

For HelloTalk, responses were very different. In general, there was much less of a consensus of the appropriateness of the language given to users in the chart below. While Very

Categories	Extremely appropriate	Moderately	Slightly	Neither	Slightly	Moderately	Extremely inappropriate
Grammar	8.33%	33.33%	0.00%	41.67%	8.33%	0.00%	8.33%
Vocabulary	8.33%	33.33%	0.00%	41.67%	8.33%	0.00%	8.33%
Written Language	0.00%	33.33%	8.33%	50.00%	0.00%	0.00%	8.33%
Spoken language/ audio	8.33%	25.00%	8.33%	50.00%	0.00%	0.00%	8.33%
Overall input (Examples of language, no what you created or produce.)	0.00%	41.67%	0.00%	41.67%	8.33%	0.00%	8.33%

Figure 18: Level appropriateness of languages given to users of HelloTalk by their language exchanges from extremely appropriate (left) to extremely inappropriate (right).

few participants rated the input from their language exchanges as outright inappropriate for their level, only 8.33% of respondents in all categories, many respondents also indicated that it was neither inappropriate nor appropriate. 50% of respondents indicated this for written and spoken language/audio, while grammar, vocabulary, and overall input were close to the middle as well

with 41.67% of respondents indicating neither appropriate nor inappropriate respectively (Figure 18).

However, expected output was rated by many participants as at least somewhat appropriate or better. Figure 19 shows that, of the participants, 41.67% found categories such as grammar and spoken language at least moderately appropriate in terms of their output, and very

Question	Extremely appropriate	Moderately appropriate	Slightly appropriate	Neither appropriate nor inappropriate	Slightly inappropriate	Moderately inappropriate	Extremely inappropriate
Grammar	0.00%	41.67%	0.00%	25.00%	8.33%	8.33%	16.67%
Vocabulary	16.67%	25.00%	8.33%	25.00%	0.00%	8.33%	16.67%
Written language	16.67%	25.00%	25.00%	25.00%	0.00%	0.00%	8.33%
Spoken language/audio	0.00%	41.67%	8.33%	33.33%	0.00%	0.00%	16.67%
Overall output (the language that you created in these exercises.)	0.00%	41.67%	16.67%	25.00%	8.33%	0.00%	8.33%

Figure 19: Level appropriateness of expected output with language exchanges in HelloTalk according to users.

few participants rated the output for their interactions as inappropriate at all.

It can be seen from the data here that a well-developed, simple level-test such as Duolingo's can be useful for users who have some previous experience. While the Duolingo level-test was not perfect by any means, as very few participants considered it not effective at all,

participants considered the language given to them by the app to be within their ability to understand and comprehend, avoiding outright frustration with language material at least.

However, from interviews with study participants, there is some evidence that suggests level appropriateness varies depending on the user's first language and the target language. Of the six interviewees, most of them were happy with the level appropriateness. Jose mentioned that he felt the instruction for French, a language that he learned from scratch on the app, was appropriate: “Yes, [I think it’s appropriate]. Even for a very beginner like I was, it was not very hard. I think they [the writing and grammar lessons], were very appropriate.”, and this sentiment was echoed in other interviews by half of the interviewees. However, other interviewees disagreed, especially if they were studying something outside of western languages. Sarah, one of the users studying Korean argued that the lessons were completely outside of her skill level (Appendix I). One respondent in particular, Christopher, who had used Duolingo for several languages to help with travelling, was able to comment on both western and eastern languages on the app: “Uhm. . . I’d say that the German, the German course was generally, in terms of skill level, was easy to manage whereas with the Asian languages, for example, I had no chance, [I was] in way over my head.”. Because these differences in perceived appropriateness are among the lines of languages that use the Romanized alphabet and languages that have different writing systems, these differences in perceived appropriateness may be associated with instructional materials around the writing systems, or lack thereof. This will be investigated in more detail in the next section about focus on form.

The results for HelloTalk, in comparison to Duolingo, are much less clear however, with many participants seemingly dissatisfied with the input while considering the output to be appropriate for their abilities. Additionally, while the accuracy of self-reporting has been shown

to be somewhat accurate for people in certain academic circles (Reuland, 2009), the self-assessment accuracy for the general population is unknown, and it is also possible that participants may be unsure of their levels. At this stage of the research, it is unclear if the unfavorable results are because of the self-reporting nature of the application, or if this is due to other factors. Additional data in the focus on form and interaction segments of this study will reveal some of the reasons behind the responses of perceived appropriateness.

4) Focus on Form

In the lessons themselves, much of the feedback in Duolingo seems to rely on written recasts. These recasts appear at the bottom of the screen when the question has been answered



Figure 20: Corrective feedback when an answer is wrong.

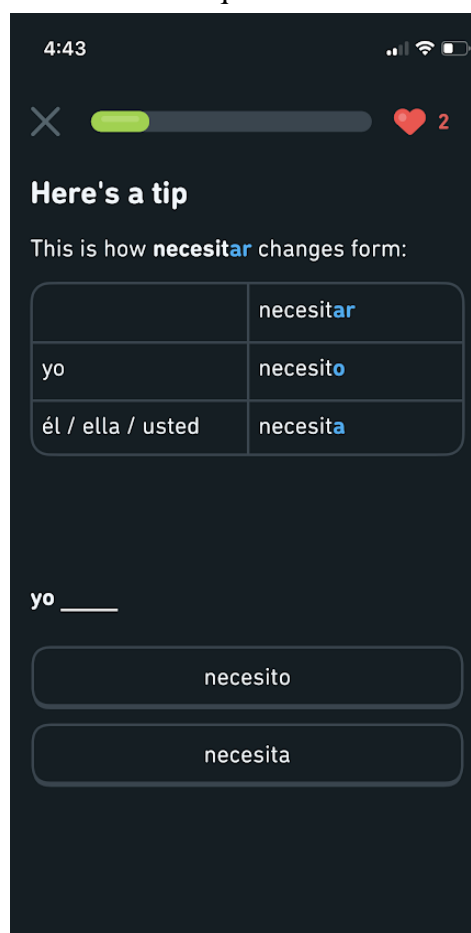


Figure 21: Conjugation explanations if user chooses the "show tip" function.

incorrectly. For all types of errors, a simple recast is the most common style of feedback. This includes verbal and listening exercises, where the desired sentence that was recorded by the program or learner is written down at the bottom of the page. With common grammar errors, sometimes learners are prompted with an option to see an explanation underneath the recast, which results in being taken to an exercise that explains the verbs and how they conjugate. Figure 20 shows an illustration of what the corrective feedback looks like. It can be seen that the learner in the example answered the question with the wrong present-tense conjugation, and the corrective recast on the bottom shows what the desired answer is. Because this is considered a common mistake, the option to see a tip appears on the bottom, but this is not given for all errors. Figure 21 is a sample of what the learner sees if they decide to look at the tip. In this case, the tip provides a simple conjugation table for present tense, with the infinitive of the verb, the form that the learner got wrong, and the desired form, showing the connection between the pronoun and the verb endings.

The app design of HelloTalk is much more focused on interactions between users rather than a developed set of lessons. However, HelloTalk still has lessons and materials available to learners who wish to have a more organized language learning experience depending on the language the learner wishes to study. Because these materials are not consistently available across languages, and because they are behind a paywall, they were not investigated in the app overview and were not asked about in the survey.

There are still functions on the free version of the app to make meetings productive in terms of Focus on Form. The application has a function for making corrections to a language

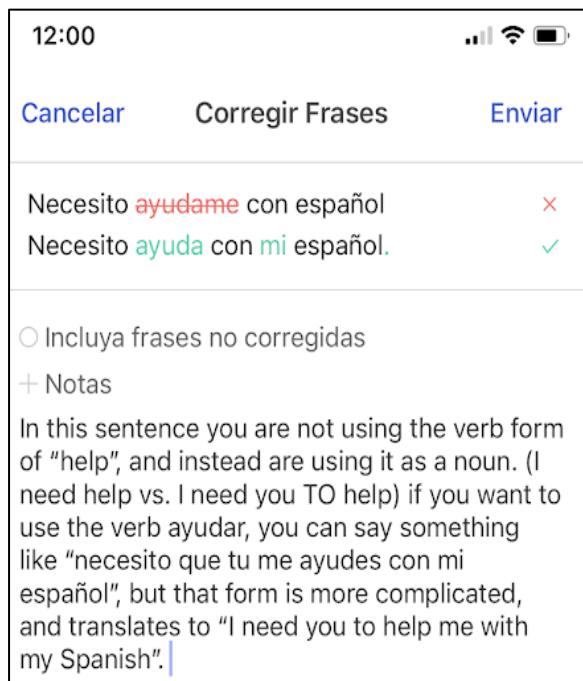


Figure 22: Sample of correction & notes function on HelloTalk.

exchange's written sentences. These corrections and explanations can be as detailed or simple as the language exchange decides they want it to be. As can be seen in figure 22, with the right language exchange, there is potential for detailed feedback on language use. Only the survey responses will give a hint as to whether the feedback from language exchanges is useful for users on average though.

In the case of Duolingo, the overall effectiveness of its style of grammar instruction was not agreed upon by participants but trended

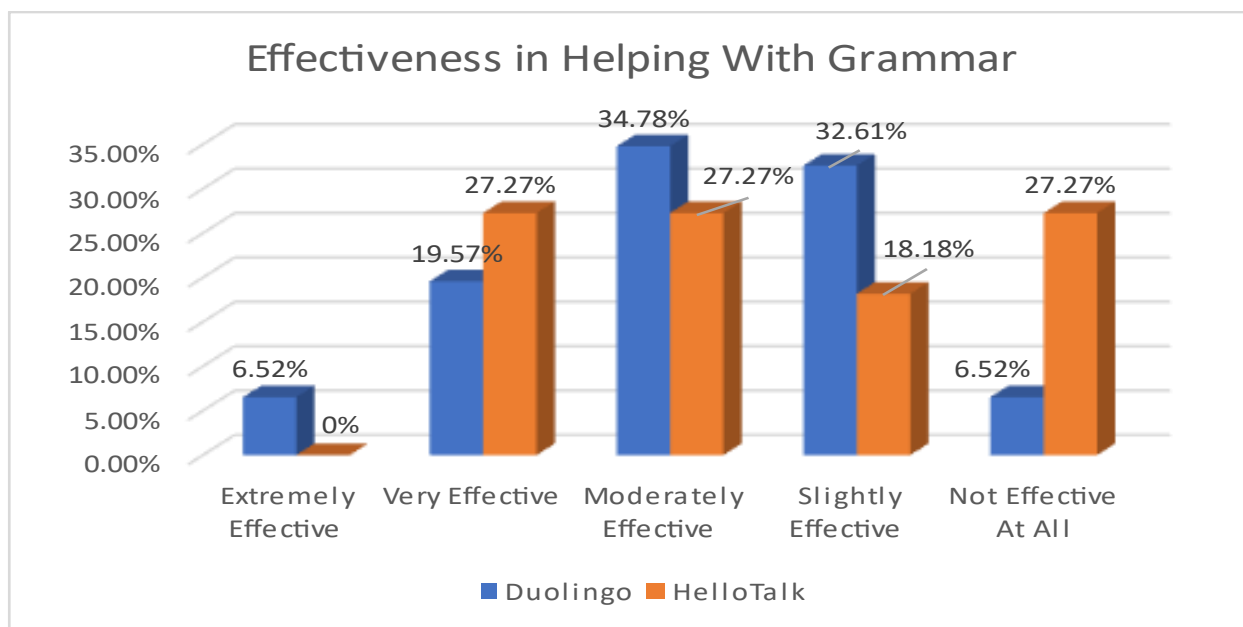


Figure 23: Effectiveness in helping with Grammar according to users. Comparison between Duolingo and HelloTalk.

towards moderately effective. While a simple majority of respondents, 34.7% of respondents, replied that it was moderately effective, 32.61% of participants stated that it was slightly effective and 19.57% of respondents considered the grammar instruction to be very effective. 6.52% of respondents considered the instructional materials extremely effective and not effective at all. In terms of how users felt about the Grammar assistance, HelloTalk has a more consistent distribution of opinions as shown in figure 23. While no one claimed that HelloTalk was extremely effective, an almost equal portion of participants claimed it to be Very Effective, Moderately Effective, and Not effective at all, while a smaller percentage said that it was slightly effective in helping them with grammar.

The answers to the short-answer question were just as mixed for both applications. Many respondents said in some way or another, when referring to Duolingo, that the grammar instruction was helpful for learning the basics or foundational grammar, but a few expressed frustrations at the oversimplification of the grammar instruction (Appendix B). Two of the best articulated complaints were that “It gives you grammar piecemeal, without the larger language context” and that “The grammar was over-generalized and didn't go into depth.”. These comments hint towards a minor issue that may be a root cause of the mediocre ratings that participants gave Duolingo: oversimplified grammar explanations. When commenting on Duolingo’s grammatical instruction in-interview, Jose similarly mentioned that the app was useful for “. . . foundational things for you to start building your sentences.”, but for more complex concepts he often “. . . had to go online and kind of do a little research to understand. . .”. This comment corroborates some other neutral to negative short-answer responses where participants mention that they get frustrated with the grammar instruction and do not understand the details of how the language works in many cases (Appendix B). This suggests that Duolingo

might be useful for a foundational understanding of target language grammar, but at some point, the usefulness of the grammar instruction drops off as students get into the higher skill levels and need to use more complex grammar forms. While this might not cause direct frustration leading to attrition rates, it is possible that many of the nearly 60% of users who ceased to use the app did so because they plateaued or thought that the app no longer helped them at their level.

Many interview participants also hinted at a problem with the pronunciation instruction in Duolingo, with the primary complaint being that it often fails to register valid language, and other times it will register complete gibberish. Both Stephanie and Mary mentioned that, for French, the app is difficult with pronunciation, with Mary explaining her frustration with it very clearly: “Sometimes I can do it perfectly, and it’ll mark me as wrong. I could be doing it incorrectly and it’ll be like ‘good job!’, so I think that’s the biggest flaw that I can point out with Duolingo; the speech recognition function.”. This comment would suggest a flaw with the software itself, but another Interviewee suggested that, in her case, it was due to the app being too strict: “I think it was a little strict [focus on pronunciation], because if I spoke Korean back in Korea people could understand me and didn’t necessarily call me out on it like, AHH! oh you’re speaking is awful, but I could also shape up in pronouncing my words a little better.” (Sarah) While it is not completely clear if this perception of the pronunciation exercises is because of the app being strict in terms of pronunciation or if the program itself is flawed, such issues can contribute to the frustration issues and attrition rates.

For HelloTalk, the responses to the short answer question sheds some light on the variety of ratings here. One respondent was extremely happy with the feedback that they received, stating that their language exchanges “. . . provided [feedback that was] extremely contextually appropriate and up-to-date. If there ever was an issue with a grammar point, they mentioned it

immediately and provided another example to further explain.”, proving that it is possible to receive great feedback with the application as it is. However, much of the dissatisfaction, according to respondents, seems to be issues with the community. One respondent stated that “The other users didn't provide much helpful correction. . .”, and this was echoed by four other short-answers to the question (Appendix F) being discussed as well as Christopher in interview who admitted that the correction function “. . . feels very helpful because you can explain why they've gone wrong, you can give them a different sentence. . .“ but laments later in the interview that he rarely found people who were willing to help him in such a way. Ultimately, user experience and perception of usefulness here seems to have completely depended on whether the users found good language exchanges. These issues were similarly echoed in questions about vocabulary and pronunciation.

The differences in responses to both the scale questions and short answer questions are likely related to the teaching philosophies of these applications. Duolingo, for their most used languages, has lessons that are fully developed and fleshed out, and while users on average do not claim that Duolingo is extremely effective, the ratings for the application are still fairly consistent towards moderately effective. The value of this level of planning and design is shown by how few participants indicated that they did not consider the material effective at all, even if the grammar is a little too oversimplified to account for the levels of Focus on Form needed for L2 learners as suggested by Gass and Selinker (2008). However, it is clear that HelloTalk is a much more hit or miss experience for users, depending entirely on if they meet someone on the app that is willing to help them learn and is willing to give them detailed feedback on forms.

5) ZPD and Scaffolding

Duolingo does have a mechanism in place for scaffolding. Most of the lessons are completed in five stages, with the language used becoming slightly more complicated as the stages advance. Additionally, shared themes are repeated later down the chain of units, and these new lessons on the same topic introduce new vocabulary and more complex, and relevant, grammar structures useful in that context. Since HelloTalk's free version depends on interaction with language exchanges, there are not any developed lessons available, and therefore scaffolding was not present in the app and not asked about in the survey.

The sentiment of participants towards Duolingo's scaffolding methods is lukewarm in general. Figure 24 is a breakdown of participants' opinions towards Duolingo's effectiveness at

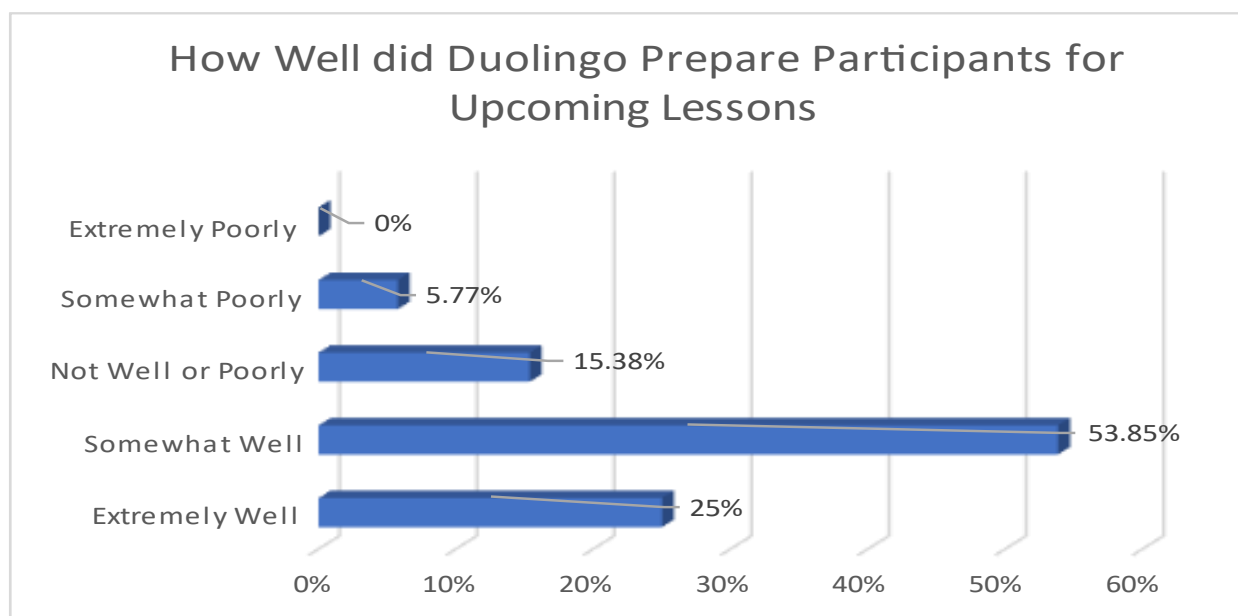


Figure 24: How well Duolingo prepared users for future lessons according to users.

preparing students for future lessons. Most participants, 53.85%, said that they felt the lessons prepared them for future material at least somewhat well, and 25% said that Duolingo prepared them for future lessons extremely well. Only 5.77% of participants stated that the instructional

materials prepared them for future lessons poorly. The responses to the questions about Duolingo lessons reviewing materials is a little more mixed, however. 36.54% of respondents felt that new lessons reviewed previous material adequately and 46.15% of respondents felt that Duolingo's lessons reviewed materials only somewhat adequately, while 17.31% of respondents stated that it did not review materials adequately at all.

These findings are consistent with the related findings on Duolingo about level appropriateness. It would make sense that if Duolingo goes through the trouble to incorporate level appropriateness, then the materials mentioned above would be associated with a certain skill level and also be intended to prepare participants for future materials as well. When asked about this aspect of Duolingo, Jose mentioned that "I think it's pretty good at building up a foundation, because it locks lessons so you can't just skip to the final lesson.". This shows that the application is implementing some level of control as to where learners are and where they are going in the program. Thus, while the implementation is not perfect, Vygotsky's concepts of ZPD and scaffolding are present in Duolingo's application, and the application is trying to build on previous materials and that users recognize that as well.

6) Interaction

Duolingo's program does not have any chances for students to interact with real people outside of forums that are more about discussing the material in the lessons rather than using the target language to interact with peers. Thus, this section will primarily focus on HelloTalk.

As a language learning app, HelloTalk is supposed to be almost entirely driven by peer interaction. Through responses to the qualitative questions and the short answer responses, the quality of these interactions seems to vary widely depending on who the learner meets and

whether they use the forum function. As mentioned previously, for both the discussion on vocabulary and grammar, some respondents reported that the interactions were helpful and informative, but others reported being unable to find anyone who was willing to give useful feedback (Appendix E, Appendix F).

This issue was made especially apparent in the interviews, where every interviewee who used HelloTalk was not really satisfied with people they met on the app and its effect on their language ability. In a previous section, Christopher's frustration in the app was made clear in his comments about focus on form, but he is not alone in that frustration. Michael mentions in his interview that many users tended to not take the language exchange part seriously or would use it as a way of finding dates. Mary's comments were particularly enlightening as she got flustered and frustrated just talking about it: "[The interactions] were really helpful, even if they were simple, because I would know and eventually, what I knew by writing down, ya know, what I knew would gradually, the pool would get bigger as I kept doing it. I just. . . (Deep sigh) no one would give me that opportunity.". While HelloTalk gives the option to interact with peers, it seems from these testimonials that making the interactions useful or meaningful in any way for language learning is difficult without some sense of direction. When participants did find helpful language partners, the app functioned well. However, many could not find a helpful language partner or did not good interactions out of the partners they found.

At this point in the analysis, it is becoming apparent that either HelloTalk's design, community, or language learning philosophy did not work for a significant portion of participants. While it cannot be stated definitively what the issue is, many of the interviewees and respondents to the short answer questions cite issues with the community at large and a lack of focus, which may be a deep flaw in a reliance on unguided interaction. With an app that relies

on the interactions of its community to achieve the primary goal, the end results are entirely unpredictable, and while some participants find good language partners and stick with it, many do not. While peer interaction as discussed by Lowen and Sato (2018) may be beneficial for language learning, a large portion of the data here is hinting that, in a smartphone context, it may not work in a vacuum.

7) Communicative Language

Duolingo contains transparent methods for attempting to teach useful communicative language for real-life situations. First, the vocabulary and grammar are tailored to specific contexts in each lesson, so that users can build vocabulary and grammar in-context during exercises. These contexts are displayed in the name of the lesson displayed to users. It should also be mentioned in this aspect that Duolingo offers podcasts so learners can listen to the language actually being used as well as *Duolingo Stories*, shown to the right, which is a program that provides conversational language in a story-like context for students to do exercises for comprehension, grammar, and vocabulary. Figure 25 showcases how the Duolingo stories work at the lower levels; there is a simple conversation with a comprehension question afterwards. These

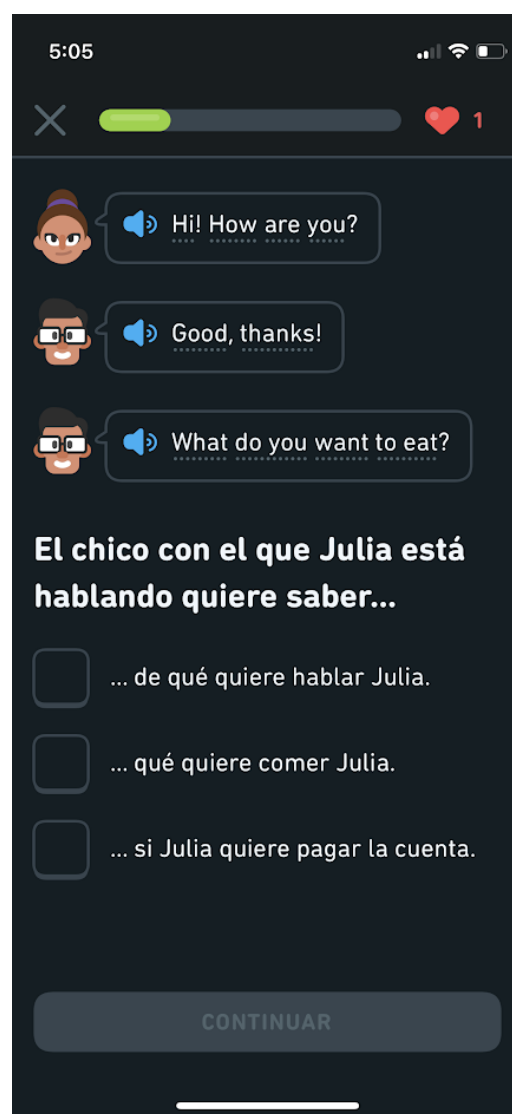


Figure 25: Sample of Duolingo Stories.

conversations get more complex overtime, but in the limitations of this review the later stages and levels were not reached.

Some users mentioned this function to be useful. Mary, in particular, really enjoys the stories function, stating that “They are simplified and streamlined for someone learning the language to understand, but it gives you an idea of ‘okay, this is how someone would respond to that question or comment’ which is, again, really useful for learning casual conversation.”, showing that at a certain level it does help with potential conversations. On the other end of the spectrum, Jose did not find them very useful, and argued that sections contained language that was way too oversimplified for casual conversation: ““I think they [Duolingo stories] get a little too dumbed down sometimes. . . “. In this comparison, it is helpful to note that Jose is at a much higher level in his target language than Mary and tends to interact with the target language in casual conversation with friends while Mary is more concerned about performing basic daily functions. These differences in level and intended uses of the languages likely affect the reasons they like or dislike Duolingo stories for helping with conversation.

For grammar in context, the results from Duolingo respondents were mixed. While very few users considered the program useful in the context of writing in general or for formal speeches, the distribution for formal conversation is spread out between very well and not very well at all. The distribution for casual conversation is even more spread out. As can be seen in figure 26, the trend for all aspects of language in context was towards slightly well and not well

Question	Extremely well	Very well	Moderately well	Slightly well	Not well at all
Speaking: casual conversation	7.69%	21.15%	26.92%	19.23%	25.00%
Speaking: formal conversation	5.77%	11.54%	23.08%	32.69%	26.92%
Speaking: speeches	3.85%	1.92%	17.31%	30.77%	46.15%

Figure 26: Effectiveness of grammar instruction in context for Duolingo according to Users.

Writing: emails and notes	7.84%	11.76%	27.45%	23.53%	29.41%
Writing: formal essays and articles	1.92%	5.77%	9.62%	30.77%	51.92%

Figure 26 (cont.)

at all, except for casual conversation and formal conversation where distribution was more spread out., where 7.69% of respondents said that Duolingo did very well at helping with grammar in context, 21.15% said very well, 26.92% said moderately well, 19.23% said slightly well, and 25% felt it did not do well at all in helping with casual conversation. Writing emails and notes also fared well, with 46.15% of participants saying that Duolingo helped at least moderately.

In comparison, HelloTalk does not have any specific features for instilling communicative language except for the language exchanges themselves. HelloTalk respondents felt considerably worse about how well the app and their language exchanges improved their grammar for specific contexts shown in figure 27. Not well at all was the majority response for all contexts, similar to the questions about grammar in context. Some respondents indicated that

Question	Extremely well	Very well	Moderately well	Slightly well	Not well at all
Speaking: casual conversation	16.67%	16.67%	25.00%	8.33%	33.33%
Speaking: formal conversation	0.00%	8.33%	16.67%	16.67%	58.33%
Speaking: speeches	0.00%	8.33%	25.00%	0.00%	66.67%
Writing: emails and notes	0.00%	8.33%	33.33%	8.33%	50.00%
Writing: formal essays and articles	0.00%	0.00%	8.33%	25.00%	66.67%

Figure 27: Effectiveness of language exchanges helping with grammar in context in HelloTalk, according to users.

they found the interactions helpful for casual conversation, but most respondents considered it not good for many other aspects of communication.

For vocabulary, respondents widely praised Duolingo. In terms of effectiveness in

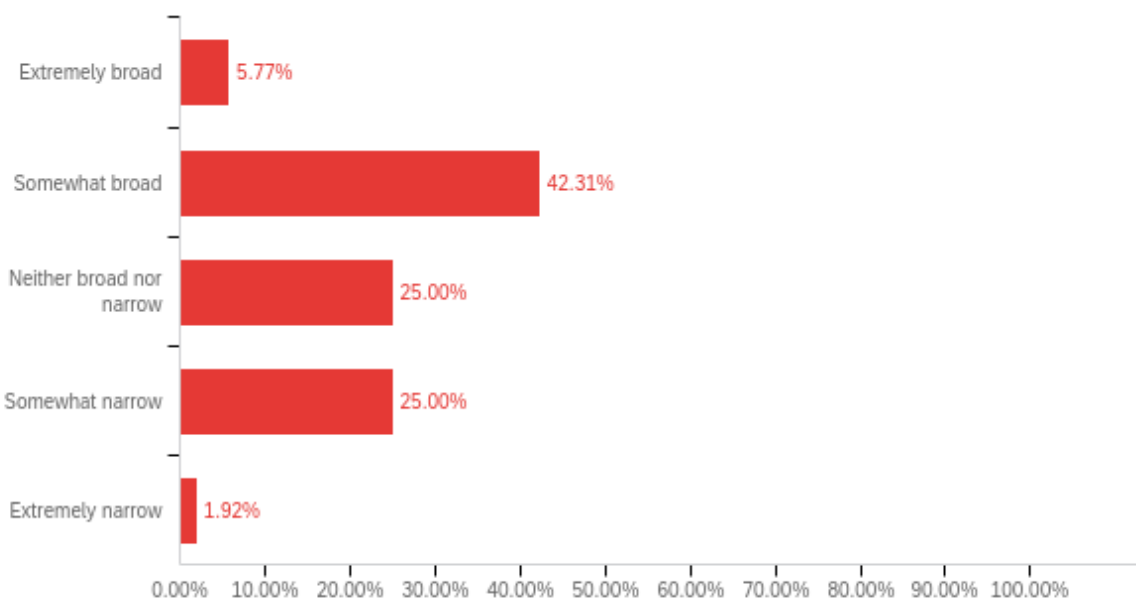


Figure 28: Range of vocabulary of Duolingo lessons according to users

teaching vocabulary in context, 78.85% of respondents who answered the question stated that the vocabulary instruction in context was somewhat effective or better, with 21.15% saying that the

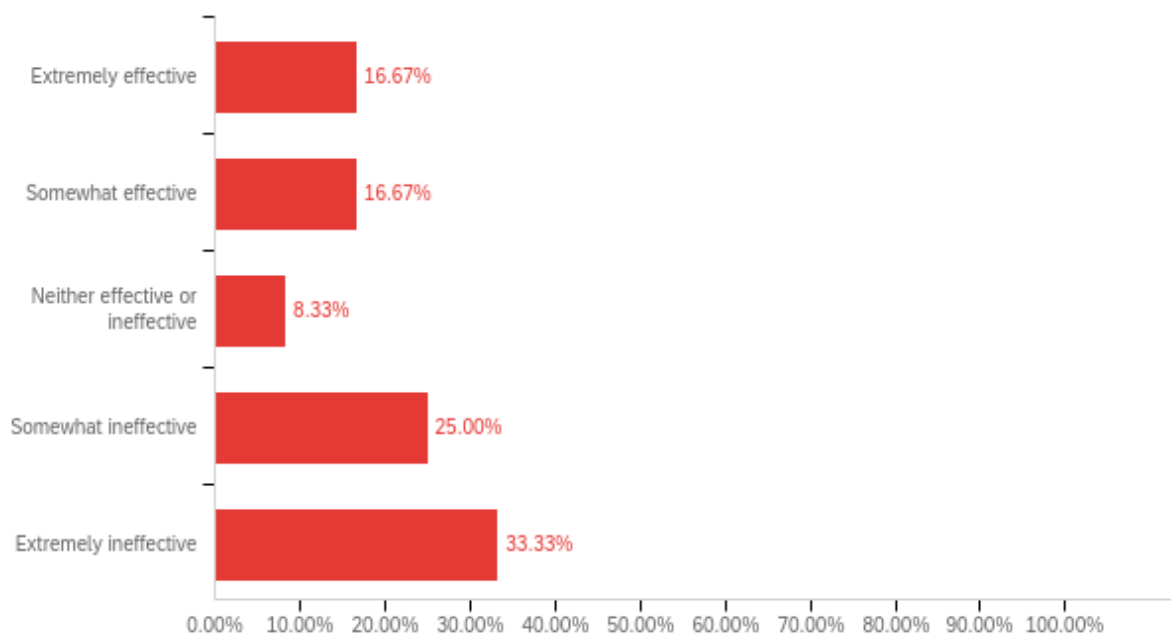


Figure 29: Effectiveness of vocabulary assistance from language exchanges according to HelloTalk users.

instruction was neither effective nor ineffective or worse. On the range of vocabulary shown in

figure 28, opinions were more mixed, with many saying that the range of vocabulary was somewhat broad, but an equal number of participants said that it was neither broad nor narrow or somewhat narrow.

HelloTalk's responses regarding the effectiveness of vocabulary in context was better than perceptions of the overall effectiveness of grammar, but still generally lower than what participants rated Duolingo. Figure 29 shows that opinions among the participants are well split but tend towards the opinion of ineffectiveness when it comes to grammar in context.

Many of the respondents to the short answer question in Duolingo followed up with promising comments discussing the fact that many of the lessons are context specific, such as lessons about restaurants and travel. One respondent explains it quite well, in stating that “[she thinks] that vocabulary learnt through Duolingo is effective in that it focuses on specific categories at a time, and it helps you learn specific words at a time.”. However, in an echo of findings by Lowen (2019), There were some complaints about frustration that the application flagged valid responses as incorrect as well most notable in pronunciation exercises, with no explanation, causing frustration (Appendix D). This would go against the concept of communicative competence, as flagging understandable sentences as incorrect is counterproductive. Additional sources of frustration reported by participants included the repetitiveness of the application and the narrowness of the vocabulary that was taught.

While the topic of communicative grammar is not all that promising for either application, it should be noted that Duolingo excels in vocabulary in context in the eyes of its users. Of course, not all users loved Duolingo's grammar instruction and grammar in context, but users consistently rated it at least moderately well, indicating that Duolingo is doing something right in this aspect of implementation. That is not to say that the application could not be

improved, in interviews both Jose and Mary expressed a desire for an in-app dictionary that can be used to see more details about words and conjugations.

For HelloTalk, it seems that grammar and especially vocabulary in context is another case of depending on user language exchanges, with the main takeaway here being that users do not find the built-in flashcard program that the app offers very useful, or do not really seem to know about it, and are thus reliant upon the language exchanges they meet. At this point, this is a recurring theme for HelloTalk, that an over-reliance on interaction with strangers on the app is resulting in very varied opinions.

8) Noticing

The concept of noticing for both applications is something that would be embedded in the focus on form instruction discussed earlier. To find data on this, this section almost entirely relies on the responses of participants in the survey, through questions attempting to ascertain if participants noticed their own errors in the application with follow up questions in-interview. However, there were some relevant survey questions in relation to this topic.

There was a consensus among participants that Duolingo was at least somewhat helpful or better at showing mistakes in grammar and vocabulary use. 55.77%, 48.08%, and 36.54% of participants stated that Duolingo was somewhat helpful in noticing grammatical, vocabulary, and pronunciation mistakes respectively, and 46.15% of participants stated that it was extremely helpful with recognizing vocabulary mistakes. However, many participants felt that the materials with pronunciation were not helpful at all, with 44.23% of participants stating so. comparatively few participants, 19.23%, stated that the pronunciation exercises were extremely helpful in terms of noticing errors. Responses to the same question in HelloTalk were fairly even across the board

with the exception of pronunciation having a high rating of not helpful at all in terms of noticing errors, with 50% of participants stating that it was not.

In the next question regarding noticing, a direct comparison between Duolingo's and

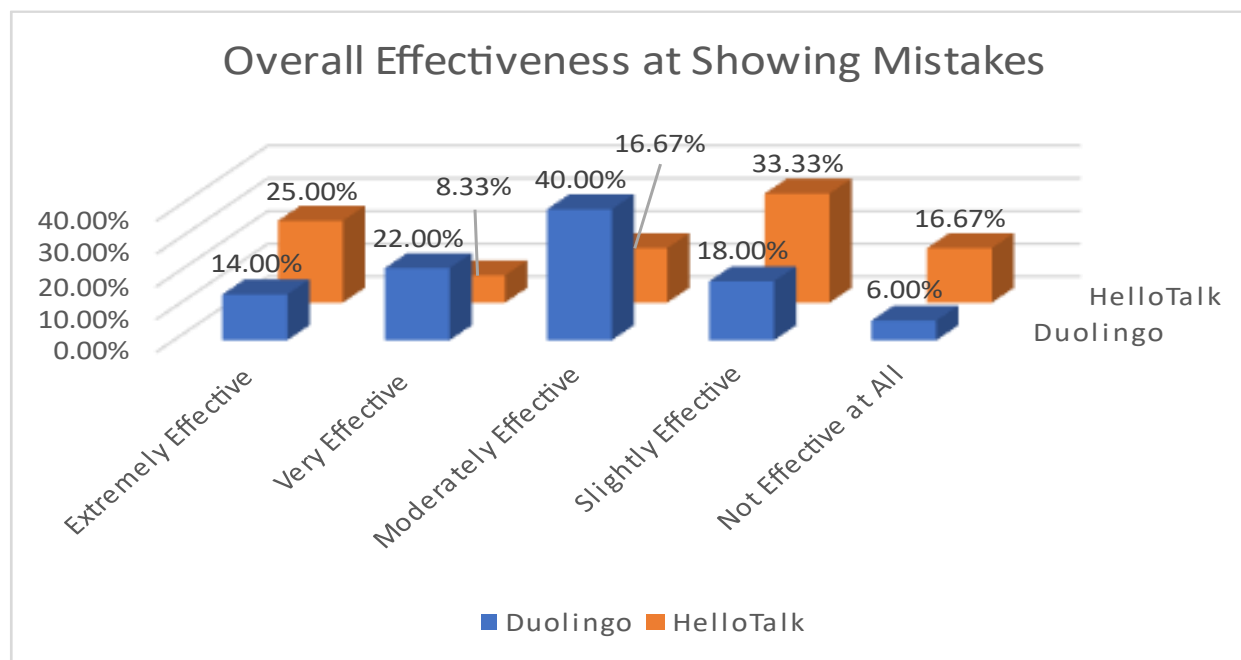


Figure 30: Overall effectiveness of both Duolingo and HelloTalk at showing mistakes according to users.

HelloTalk's results yields interesting results. From figure 30, it can be seen that opinions about Duolingo and the overall effectiveness of showing mistakes is more consistent than HelloTalk, but HelloTalk has a higher percentage of respondents who said that the feedback was extremely effective.

Some of the short-answer responses in this case help to shed some light on the responses regarding HelloTalk. For those who reported positively in the short-answer question, many of them liked that they received corrections for things that sounded natural. As one respondent put it, "[she] liked that [she] could have native speakers correct [her] responses directly and tell [her] what sounds the most natural.". One respondent even mentioned that the grammar corrections on the forums were extremely useful for helping them notice their errors, and this was echoed by

another participant, but she did mention afterward that the corrections in the individual chats were not always as helpful or as specific (Appendix E). Most of the negative comments were related to an inability to find people who were willing to give feedback in the first place similar to concerns expressed in interviews mentioned in the previous sections (Appendix F).

Reminiscent of the discussions of the previous sections where direct comparisons could be made, Duolingo is proving to be more consistent in how useful users find it for noticing as well. While some participants found HelloTalk to be really effective, Duolingo trends towards moderately effective while many participants for HelloTalk say that the application was only slightly effective or not effective at all for helping them notice their mistakes.

With the recognition that noticing has a positive impact on learning outcomes (Mackey, 2006), and that it is directly related to the forms being focused on (Kartchava and Ammar, 2014), it makes sense that Duolingo was rated rather consistently, as it has established lessons that focus on specific forms in that lesson. The variation in responses for HelloTalk are also very understandable with the previous sections of this analysis in mind, due to the unpredictable and unguided nature of the language exchanges.

9) Motivation and Gamification

Duolingo makes significant and transparent use of gamification. The first aspect of it worth mentioning is the hearts that can be seen on the top right corner of the screen when using the app that are lost when a question is answered wrong, not dissimilar to a “life” in a video game. While losing all your hearts does not end in a complete restart or a backtrack in progress like a video game, it does remove the learner from their lesson and requires them to take a practice lesson in order to regain hearts, and also gives them the option of watching an ad to

regain an additional heart. Additionally, you can acquire what the application calls gems when completing lessons, which can be used to buy things like skins for the Duolingo bird, as well as additional content and perks. Finally, Duolingo’s gamification also utilizes leader boards and rankings. Duolingo applies learners to “leagues”, similar to competitive online games, and then places learners on a scale in those leagues.

The extrinsic motivations for users of Duolingo varied, but by far, the biggest reason that participants used Duolingo to learn a language was out of personal interest. Figure 31 shows

Categories	Extremely important	Very important	Moderately important	Slightly important	Not at all important
Work	7.69%	11.54%	13.46%	21.15%	46.15%
School	9.62%	9.62%	13.46%	19.23%	48.08%
Personal interest in English or languages in general	46.15%	36.54%	13.46%	3.85%	0.00%
Communicating with friends or family	11.76%	9.80%	13.73%	13.73%	50.98%
Being able to participate in online interaction or gaming	5.77%	13.46%	15.38%	26.92%	38.46%
Travel	26.92%	19.23%	13.46%	19.23%	21.15%

Figure 31: What is important to Duolingo users.

the percentages for each aspect of motivation. 46.15% of participants marked personal interest in languages as extremely important for their motivation to study when using Duolingo, and 36.54% of participants marked it as very important. Travel was also quite important for many of the participants, with 26.92% of participants marking the category as extremely important and 19.23% of participants marking it as very important. While some participants marked communicating with friends and family as extremely important and very important, it was the weakest category with 50.98% of participants marking it as not at all important. Other similarly weak categories for outside motivation include school and work. Overall, 50% participants were

able to stay motivated to use Duolingo to practice for longer than three months, with 19.23% of respondents reporting that they were somewhat motivated after three months and 30.77% of respondents stating that they lost motivation.

As HelloTalk does not contain many of the gamification aspects to it as Duolingo, long-term motivation for using HelloTalk will come from different factors. However, it is telling that, in the end, most of the users, 66.67%, who responded to the survey no longer use HelloTalk, citing issues with being unable to find good language partners, the interactions not being helpful, or finding better modes of interaction in the associated short-answer question. One participant who responded to the short-answer question stated that she was able to find a good language exchange and stick with the app to learn, but it seems like that is the exception rather than the rule (Appendix H). It is not entirely certain if a motivational mechanism would help HelloTalk retain users here, as many of the issues with this sample group seem to be broader issues with the community rather than problems with the app design itself.

In the same short answer question for Duolingo, a few respondents mentioned gamification as a reason for continuing to use the application. One user commented that “It is like a game, and I want to play It every day”. Another user goes further and specifically mentions the streak boards “The Duolingo sub in Reddit has people posting their day streaks, and I wanna do that too”. Additional comments mention competition with people that they know. one user mentions that she “. . .really like[s] keeping the streak alive and [she] compete[s] with [her] boyfriend on building long streaks on Duolingo.” These comments are promising for Duolingo’s model of gamification, as it shows that the competition is something that is keeping a good portion of the users there.

The interviews were also enlightening in discussions about Duolingo's gamification, as this was probably one of the aspects of the apps that interviewees had the most opinions on. A few users really liked the gamification of the applications, Mary in particular. Mary discussed the gamification aspect a lot as a significant element of her decision to use the application on a daily basis:

“There is an aspect of the app, uhm, it has a streak function that, uhh, you get rewarded for keeping up a streak for so long. There's, like, no monetary reward, it's just the glory of 'you've been practicing for this many days! . . . it notices that you are putting in the effort and you do get little achievements, there's a scoreboard, that says, oh you've, you've been practicing for this long. Congratulations! . . . I have a 132 day streak and I will get a level eight achievement wildfire if I reach a 180 day streak, for example, so it's, again, those tiny little gifts they give you to kind of egg you to keep going which, manipulative in a sense yes, but in a good way!”

To add to this, both Stephanie and Sarah had comments about the competitive side of Duolingo's gamification, both speaking of it highly. For Stephanie, the competition “. . .kept [her] motivated because [she] kept being number one in the scoreboard thing and [she] was like, 'gotta keep my title!' So [she] kept playing it again and again and kept using it.”. Sarah's sentiments about the competition were similar, in the end stating that “. . . competition kept [her] on the app, mostly.”. These all corroborate the positive sentiments towards the gamification in the short-answer questions. While these users all have their own reasons for starting the app, it is clear that a big factor in the continued use of the app is the gamification.

There is an opposite side to Duolingo's gamification however, which seems to be user frustration or not taking the activities seriously because it is like a game. One frustrated user in

the short answer responses reported that they “. . . got annoyed at how much time the app spent trying to convince [them] to keep using it.”, potentially referring to the messages about losing day-streaks that Duolingo sends out occasionally. Another user also reported that “. . . it can feel like a grind. . . “, which, in a gaming context, refers to doing repetitive mindless actions to achieve an end or gain some petty reward. Jose made similar comments towards the competition being a little too much in his interview:

I can see more people getting hung up on the competition side of it. They have, like, they have a scoreboard. I can see people getting hung up. I kind of, I kind of, like when I first learned Spanish I tried a little bit, and I can see how people can get addicted to it. . . I would do one lesson and get, like, 20-30 points and then I'd see some people getting 300 points to 400 points a day and think to myself ‘don’t these people have anything else to do? So, I can see how people get kind of hung up on that. Yep, so I quit looking at the scoreboard and I just follow my own personal schedule now.”

While these comments are not the majority of opinions, there are enough negative responses about Duolingo’s gamification to show that the design can occasionally backfire or even be a non-factor with many of its users. Overall, it seems that Duolingo’s gamification helped learners stay motivated, or they did not mention Duolingo’s gamification at all, and instead mentioned personal motivation or the ease of access and user friendliness.

While external motivation is what generally brings users to the apps, it seems like effective gamification can be a good tool in keeping the attention of users over long periods of time. In HelloTalk’s case, it is unclear if the lack of gamification is the reason for user attrition or if it is related with deeper issues in the community and reliance on language exchanges for all of the learning. What is clear is, while both applications are suffering the app frustration that was

documented Loewen's (2019) study, HelloTalk seems to be frustrating its users at a higher rate than Duolingo judging by its attrition rate.

CHAPTER 6: CONCLUSION

In this study, the influence of SLA principles in language learning app design was investigated, followed by an evaluation of efficacy using user responses through a survey and interviews. This section will contain a summary of findings based on the analysis section of this study, these findings' implications for both app developers and users, limitations of the study, and suggestions for future research.

1) Summary of findings

With all the different responses and disparate data related to the SLA principles that were researched here, there was one constant throughout the study, and that is that Duolingo was rated rather consistently in the middle while HelloTalk was often rated as either extremely effective or not effective at all, with much fewer in-between ratings. The reasons for this varied depending on the SLA principle being investigated, but in general the ratings of HelloTalk often depended on whether the participants were able to find a reliable, serious language exchange or not. Meanwhile, participants considered Duolingo at least somewhat useful or better in most aspects, and especially shined in teaching vocabulary in context.

There are a few highlights of the research that showcase the differences between user perceptions of these two apps. The first of these is the Focus on Form section, where figure 23 showed the trend of Duolingo's effectiveness according to its users tends towards the middle while HelloTalk has more users reporting both very effective in terms of grammar assistance and not effective at all. From testimonials it was clear that Duolingo's users generally believed that, while the application was good for building simple grammar structures and foundational sentence building, some of the grammar instruction on the app was over-simplified, and there were significant issues with the pronunciation aspects of the lessons. Meanwhile, HelloTalk's short-answer survey responses and discussions in the interviews hinted that user experiences

were very hit-or-miss, depending on who the user met and whether they were helpful in the target language.

The second highlight that demonstrates the differences in how users felt about the applications is the communicative language segment, where Duolingo shined in user's perceptions of vocabulary for specific scenarios and in specific context, with over 78.85% of participants saying that the vocabulary instruction for specific scenarios was at least somewhat effective or better on a scale from not effective at all to extremely effective. In comparison, HelloTalk continued to get scattered ratings from its users, and evidence from interviews suggests that the reason for this is similar to the ratings for Focus on Form as well; good language exchanges are hard to find, and few users find a truly helpful language exchange to talk with them.

A final segment that really drives home these differences between users' feelings towards the app is the noticing section. In this, Duolingo once again trended towards the middle in how users felt that the app assisted them in noticing their mistakes, while HelloTalk's ratings were distributed across the ratings similar to previous segments of research. Once again, interviews and short answer responses revealed that the primary difference between an excellent experience in this regard and a poor experience relied upon finding a good language exchange that was willing to give good feedback.

A final segment that is worth highlighting here, but that does not demonstrate the differences between the apps in the eyes of their users, is the gamification section. The reason this was chosen is because of the differences in opinion towards the gamification put in place by Duolingo. While most users used the application for a significant amount of time, the reasons differed, with some participants saying that the gamification was a primary reason for continued

use, and others citing personal reasons for continuing to use the application. Others still disliked the gamification enough that they stopped using the app altogether. While the gamification present in Duolingo can certainly keep users interested in the application, these differences in opinion also suggest that it can sometimes backfire. However, it is also telling that Duolingo retains its users at a higher rate than HelloTalk. It is not clear if this retention is because of gamification or of the previously mentioned issues with finding good language exchanges in HelloTalk.

In sum, While both Duolingo and HelloTalk had users who considered their applications useful for fulfilling the SLA principles listed in the research questions, Duolingo was much more consistent in how users rated the program; while most users did not find Duolingo to be perfect in any of the categories, most considered it at least somewhat useful or better in almost all categories that were inquired about, and Duolingo especially shined in terms of Vocabulary in Context. User experiences for HelloTalk, in comparison, are significantly more unpredictable, with some users reporting to be extremely satisfied with how well their language exchanges help them with forms, noticing their errors, and communicative language, while many others stated the app to be not effective at all due to the difficulty of finding a truly helpful language exchange.

2) Implications for App Developers

One of the primary takeaways from this research is that, for many students, there is not really a substitute for planned and organized lessons. In the case of HelloTalk, interaction alone did not seem to help many users without some form of specific instruction or goal in mind. Although Duolingo does not have a method for encouraging actual interaction between peers, the app was still more consistently rated as useful or effective on some level, and the evidence implies

that this is because Duolingo, for its primary and older languages, has established lessons that provide context, some scaffolding, and useful grammar forms for specific scenarios and situations as well as vocabulary. For a casual user looking for a bit of extra practice in their target language, or a complete newbie who wants exposure to a language, this is a good app to turn to. While it is very unlikely that the app will get you to fluency, it is capable of building a foundation for learners who are dedicated enough to keep using it or who are drawn to its gamification according to many of the interviewees and users in the survey.

The most important aspect that is in Duolingo's favor is that all of these features are *free access*, whereas many of the features that users might find useful in HelloTalk, such as established lessons, podcasts, and vocab exercises beyond the absolute basics are behind paywalls. Not only does this skew the results against HelloTalk due to the limitations of this study, which was explicit use of the free versions of the app, but many of the interviewees who used HelloTalk never even mentioned paying for the extra functions that HelloTalk offers and seemed completely unaware that there were options for lessons that existed for some languages. Mary, an interviewee who used both but only stuck with Duolingo, explained the issue of the paywall perfectly in her interview:

I feel like HelloTalk is extremely restrictive if you do not buy the subscription. I can tell you with 100% confidence that you can pick up Duolingo and you don't need a subscription to learn from it. You have more freedom with the subscription, but you'll be fine without it. [For HelloTalk], I feel like you have to pay to use the app to its absolute best and I think, really, the only free option given to you is trying to speak to people in a certain language, so I found HelloTalk to be very restrictive.

Whereas Duolingo's free app restricts users in the order in which things can be completed, and how often some aspects of the app can be used, HelloTalk's free app blocks out dedicated instructional materials entirely, which is harmful to all users but particularly harmful to lower-levels who may want some instruction to establish a baseline of language.

This does not mean that HelloTalk's concept is defunct; some participants found language exchange partners that were particularly helpful. While none of the satisfied users of HelloTalk left their emails for an interview for more details, it is still clear that effective language exchanges using such an app are possible. Hulme and Viberg (2017) argue in their meta-analysis that mobile collaborative learning can be effective if it is tied with task-based learning, and in that vein I would argue that to be useful to more users and learners, some form of established instructional materials need to be freely available, and at a lower level much of the interaction in an app would need to be guided and task-based in some way more similar to the classroom interactions that were beneficial in Loewen and Sato's (2018) study. In general, having people find their own language exchanges and determine their own learning on the app seems to be unpredictable as it stands, and leads to a lot of interactions that lack purpose.

Duolingo is already constantly improving its app and lessons with the help of its community, giving it even greater potential for collaborative learning in that community ownership of the material can underpin a good communicative environment (Hulme and Viberg, 2017). Many of the minority languages are being developed almost exclusively by the Duolingo community rather than the app developers, and the community continues to modify the lessons even for the classic languages offered on the app. To play on these strengths, Duolingo could start developing further materials to help with teaching the more complex grammar in a way that makes sense for users, as users did express frustration with the app and its method of giving

grammar instruction. It would also be possible to develop activities that utilize communication such as what HelloTalk attempts to foster, with the added benefit of being part of an app that already has other supplementary materials and guided learning. With their framework it may be possible to develop relevant interactions for scenarios utilizing meaningful interaction with the environment and/or others more akin to what Robert-Godwin Jones (2017) originally imagined.

3) Implications for App Users

It is doubtful that mobile apps will replace traditional language classes anytime soon; many of the users who participated in interviews expressed opinions that Duolingo is supplemental in nature or developed a foundational knowledge. Additionally, it is important to stay focused. The most satisfied users of Duolingo, who happen to be many of the interviewees as well, were able to consistently do exercises on a daily basis for long periods of time. Mary's earlier comment on being on a streak for longer than 130 days in a row hints at least a little bit to the amount of dedication needed to self-learn a language using mobile apps. With focus, it is possible. For potential users who like challenges and rewards, the gamification of Duolingo seems to keep many of its users engaged for long periods of time as well and may be a great way to get started.

After building a foundation, interaction of some sort, perhaps in an app like HelloTalk, should be the next step along with other learning materials, as a foundation gives you the ability to understand and respond at least on a basic level. However, as shown by the analysis in this paper, finding a good language exchange on the app may be difficult. While the specific reasons for *why* it was difficult to find good language exchanges were not explored in this study, a quick glance at the app and an attempt to communicate with people suggests that a certain level of

personal drive and motivation may be required to find someone to work with who is actually helpful.

The market of language learning apps is huge as well, and these are not the only two apps that should be looked at when deciding to study a language. In the end, the biggest determiner of the learning outcome is motivation to find a tool that works and develop your own personal learning method that works for you.

4) Limitations of This Study

The results of this study are not definitive by any means. The first, and most notable, issue of this study is the sample size of participants. In total, only around 101 respondents gave valid responses to the survey online. This already small sample size is compounded by the fact that these participants are split among two different apps. In a perfect scenario, this split would have been even, but it was not; an overwhelming majority of participants responded for Duolingo and Duolingo only. While this comparison between the two apps is food for thought, the disparity in sample sizes makes it impossible to make any definitive claims from the data and makes the data for HelloTalk even more unreliable than it already would be given the sample size of 101 total participants.

A second limitation that is directly related to the sample group is the spread among different demographics. It proved difficult to find users who spoke English as a second language and used the app for English and to get those users to attend focus groups and interviews. The reasons for this can only be speculated, but it is likely from the demographics taken in this sample that many users that are using Duolingo for English are likely to be intermediate to beginner level students in the language, and thus might have reservations about sinking time into taking a survey in English. Additionally, the survey participants that *did* learn English on

Duolingo were very unlikely to leave an email for contact for interviews in the first place, and the few that did leave emails did not respond. Ideally, future studies would focus on specific groups of learners learning a specific language, such as Spanish speakers using the app for English, and provide the survey in the L1 of the target audience to make the survey easier for participants to take. With a more concentrated group of learners in mind, more definitive results can be inferred with the same number of participants. In this study, the first language and second languages of the participants was completely open, and thus the ratings for Duolingo are across all of the possibilities that Duolingo has to offer.

Another major limitation of this study is that this analysis focused on two different apps. There are two other ways that this study could have gone that may have been more productive. First, a very focused investigation on a single app and its strengths and weaknesses could have realistically been done for Duolingo using the SLA concepts in the study as evaluation criteria. Second, expanding the study to more apps would have been more helpful for generally gauging the effectiveness of language learning apps. While it was useful that this study focused on two very different applications and philosophies for comparison, a wider sample group or a more focused study on one app may have resulted in more reliable data.

A final limitation of this study is the wide swathe of SLA principles that were covered in the study. While this made sense for research limited to two applications, another useful way of investigating these would be to dedicate a study to identifying a specific SLA theory in a multitude of apps and gathering survey data to evaluate the implementation. This would give researchers a chance to specifically target aspects of these apps individually, which would be easier for direct comparisons.

5) Suggestions for future research

For more reliable results on language apps and their function in individuals' learning journeys, more research will need to be done to make up for the limitations of the research mentioned above. A similar study to this with a significantly larger sample size needs to be done to verify user perceptions of Duolingo and also verify the overall sentiment towards HelloTalk and its interactional approach. Another modification to be made in future research that would be helpful would be to expand the number of applications being investigated so general information can be gathered about apps.

For general data on what apps are being used for, a large-scale study to determine how many language learners are utilizing these apps and what else they do outside of the apps would be needed. Some of the participants of this study, especially interviewees, hinted that they considered these apps supplemental in nature, suggesting that they do something outside of the app itself. Jose mentioned that he spends more time reading and watching TV in the target language than he does on the app after a certain level. A study to find if this is the norm or if users are solely relying on these apps to learn would be helpful in determining how detailed many of these apps need to be in their instructional materials or designs.

A final line of inquiry would be research on how to integrate language learning apps into classroom instruction. Mobile apps are likely to continue growing in popularity, so the most important thing for researchers in language learning is to determine how these apps are being used, and what students feel these apps are missing in terms of their overall learning experience. While apps are not going to replace an actual classroom and teacher any time soon, the potential for these apps to cover a lot of the more tedious material such as vocabulary memorization and

grammar drilling have the potential to make in-person classes more productive and communicative in nature.

REFERENCES

- Al-Sabbagh, K. W., Bradley, L., & Bartram, L. (2019). Mobile language learning applications for Arabic speaking migrants – a usability perspective. *Language Learning in Higher Education*, 9(1), 71–95. <https://doi.org/10.1515/cercles-2019-0004>
- Allum, P. (2004). Evaluation of CALL: Initial vocabulary learning. *ReCALL*, 16(2), 488–501. <https://doi.org/10.1017/s0958344004001624>
- Burston, J. (2014). MALL: The pedagogical challenges. *Computer Assisted Language Learning*, 27(4), 344–357. <https://doi.org/10.1080/09588221.2014.914539>
- Burston, J. (2014). The reality of MALL: Still on the fringes. *CALICO Journal*, 31(1), 103–125. <https://doi.org/10.11139/cj.31.1.103-125>
- Burston, J. (2014). Twenty years of MALL project implementation: A meta-analysis of learning outcomes. *ReCALL*, 27(1), 4–20. <https://doi.org/10.1017/s0958344014000159>
- The CEFR Levels*. Common European Framework of Reference for Languages (CEFR). (n.d.). <https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions>.
- Chang, M.-M., & Hung, H.-T. (2019). Effects of technology-enhanced language learning on second language acquisition: A meta analysis. *Educational Technology & Society*, 22(4), 1–17.
- Chang, S.-C. (2011). A contrastive study of grammar translation method and communicative approach in teaching English grammar. *English Language Teaching*, 4(2), 13. <https://doi.org/10.5539/elt.v4n2p13>

- Chapelle, C. A. (2009). The relationship between second language acquisition theory and computer-assisted language learning. *The Modern Language Journal*, 93, 741–753. <https://doi.org/10.1111/j.1540-4781.2009.00970.x>
- Chen, S., & Nassanji, H. (2018). Focus on form and corrective feedback research at the University of Victoria, Canada. *Language Teaching*, 51(2), 278–283. <https://doi.org/10.1017/s026144481800006x>
- Doughty, C., & Varela, E. (n.d.). Communicative Focus on Form.
- Garcia, I. (2013). Learning a language for free while translating the web. Does Duolingo work? *International Journal of English Linguistics*, 3(1). <https://doi.org/10.5539/ijel.v3n1p19>
- García Botero, G., & Questier, F. (2016). What students think and what they actually do in a mobile assisted language learning context: new insights for self-directed language learning in higher education. *CALL Communities and Culture – Short Papers from EUROCALL 2016*, 150–154. <https://doi.org/10.14705/rpnet.2016.eurocall2016.553>
- Godwin-Jones, R. (2011). Emerging technologies: Mobile apps for language learning. *language Learning & Technology*, 15(2), 2–11.
- Godwin-Jones, R. (2017). Smartphones and language learning. *Language Learning & Technology*, 21(2), 3–17.
- Hirschel, R., & Fritz, E. (2013). Learning vocabulary: CALL program versus vocabulary notebook. *System*, 41(3), 639–653. <https://doi.org/10.1016/j.system.2013.07.016>
- Hymes, D. H. (1986). On communicative competence. In *Sociolinguistics: Selected Readings* (pp. 269–293). essay, Penguin Books.
- Ibacache, K. (2019). Use of language-learning apps as a tool for foreign language acquisition by academic libraries employees. *Information Technology and Libraries*, 38(3), 22–33. <https://doi.org/10.6017/ital.v38i3.11077>

- Kartchava, E., & Ammar, A. (2014). The noticeability and effectiveness of corrective feedback in relation to target type. *Language Teaching Research*, 18(4), 428–452. <https://doi.org/10.1177/1362168813519373>
- Kukulska-Hulme, A., & Viberg, O. (2017). Mobile collaborative language learning: State of the art. *British Journal of Educational Technology*, 49(2), 207–218. <https://doi.org/10.1111/bjet.12580>
- Lantolf, J. P., & Aljaafreh, A. (1995). Second language learning in the zone of proximal development: A revolutionary experience. *International Journal of Educational Research*, 23(7), 619–632. [https://doi.org/10.1016/0883-0355\(96\)80441-1](https://doi.org/10.1016/0883-0355(96)80441-1)
- Loewen, S., & Sato, M. (2018). Interaction and instructed second language acquisition. *Cambridge University Press*. <https://doi.org/https://www.cambridge.org/core/journals/language-teaching/article/interaction-and-instructed-second-language-acquisition/78A156EE200F744F5978F99BFB073DBE>
- 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>
- Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics*, 27(3), 405–430. <https://doi.org/10.1093/applin/ami051>
- Nielson, K. B. (2011). Self-study with language learning software in the workplace: What happens? *Language Learning and Technology*, 15(3), 110–129.
- Petterson, E. H. (2019). *Pedagogy and Usability in Language Learning Apps* (thesis). Masters Abstracts International, Ann Arbor. Retrieved from <https://www-proquest->

com.proxy2.library.illinois.edu/dissertations-theses/pedagogy-usability-language-learning-apps/docview/2503626514/se-2?accountid=14553.

Pew Research Center. (2021, April 26). *Demographics of Mobile Device Ownership and Adoption in the United States*. Pew Research Center: Internet, Science & Tech. <https://www.pewresearch.org/internet/fact-sheet/mobile/>.

Read, T., & Bárcena, E. (2020). Hacia un marco para los MOOC de idiomas y el aprendizaje de idiomas con asistencia móvil. *Propósitos y Representaciones*, 8(1). <https://doi.org/10.20511/pyr2020.v8n1.470>

Reuland, D. S., Frasier, P. Y., Olson, M. D., Slatt, L. M., Aleman, M. A., & Fernandez, A. (2009). Accuracy of self-assessed Spanish fluency in medical students. *Teaching and Learning in Medicine*, 21(4), 305–309. <https://doi.org/10.1080/10401330903228489>

Rosell-Aguilar, F. (2016). State of the App: a taxonomy and framework for evaluating language learning mobile applications. *CALICO Journal*. <https://doi.org/10.1558/cj.27623>

Sauro, S. (2009). Computer-Mediated Corrective Feedback and the Development of L2 Grammar. *Language Learning & Technology*, 12(1), 96–120.

Silver, L. (2020, August 25). *Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally*. Pew Research Center's Global Attitudes Project. <https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/>.

Smith, A. (2020, August 25). *Usage and Attitudes Toward Smartphones*. Pew Research Center: Internet, Science & Tech. <https://www.pewresearch.org/internet/2015/04/01/chapter-two-usage-and-attitudes-toward-smartphones/>.

- Stiller, K. D., & Köster, A. (2016). Learner attrition in an advanced vocational online training: The role of computer attitude, computer anxiety, and online learning experience. *European Journal of Open, Distance and E-Learning*, 19(2), 1–14. <https://doi.org/10.1515/eurodl-2016-0004>
- Tuncay, H. (2020). *App attrition in computer-assisted language learning: focus on Duolingo* (thesis). Masters Abstracts International, Ann Arbor. Retrieved from <https://www-proquest-com.proxy2.library.illinois.edu/dissertations-theses/app-attrition-computer-assisted-language-learning/docview/2515598254/se-2?accountid=14553>
- Vygotsky L. S., Cole, M., Stein, S., & Sekula, A. (1978). *Mind in society: the development of higher psychological processes*. Harvard University Press.
- Wang, Z., Hwang, G.-J., Yin, Z., & Ma, Y. (2020). A contribution-oriented self-directed mobile learning ecology approach to improving EFL students' vocabulary retention and second language motivation. *Journal of Educational Technology & Society*, 23(1), 16–29.
- Werbach, K. (2014). (Re)defining gamification: A process approach. *Persuasive Technology*, 266–272. https://doi.org/10.1007/978-3-319-07127-5_23
- Williams, J. N. (1999). Memory, attention, and inductive learning. *Studies in Second Language Acquisition*, 21(1), 1–48. <https://doi.org/10.1017/s0272263199001011>
- Yilmaz, Y. (2012). Relative effects of explicit and implicit feedback: The role of working memory capacity and language analytic ability. *Applied Linguistics*, 34(3), 344–368. <https://doi.org/10.1093/applin/ams044>

APPENDIX A: SURVEY

Language Learning Apps

Start of Block: Permissions

1 You are being asked to participate in a research study being done by Dr. Randall Sadler at the University of Illinois at Urbana-Champaign. Your participation in this research is voluntary. If you decide to participate, you are free to withdraw at any time. The purpose of this study is to identify and evaluate the use of language learning theory in popular language learning applications, and to see how students feel about the effectiveness of these applications' use of the language learning theories. Participating in this research study will include a survey, to be possibly followed up by a recorded focus group or an interview if you give us permission to do so. The survey will take about 30 minutes to complete. The focus groups and interviews would take roughly 30 minutes to an hour to complete if you chose to participate in them, but it is not required. We will ask everyone in the follow-up focus groups to respect the privacy of other participants and to treat anything said in the group as confidential. However, please remember there is no guarantee that other participants will abide by that request. These focus groups and interviews will be done through Zoom and will be recorded for research purposes. These recordings will NOT be shared by anyone outside of the research team. If you do not wish to participate in the interviews and focus groups, but would still like to do the survey, then in question #2 select "I do not consent to being contacted". Faculty, students, and staff who may see your information will maintain confidentiality to the extent of federal and state laws and university policies. Personal identifiers will not be published or presented. Your de-identified information could be used for future research without additional informed consent. If you have any questions about the research study, please contact Dr. Randall Sadler at rsadler@illinois.edu. If you have questions or concerns about your rights as a participant please contact the University of Illinois at Urbana-Champaign Office for the Protection of Research Subjects at 217-333-2670 or via email at irb@illinois.edu. By clicking the button below, you acknowledge that your participation in the study is voluntary, you are 18 years of age, and that you are aware that you may choose to terminate your participation in the study at any time and for any reason.

- I consent, begin the study (1)
- I do not consent, I do not wish to participate (2)

Skip To: End of Survey If You are being asked to participate in a research study being done by Dr. Randall Sadler at the Un... = I do not consent, I do not wish to participate

2 (optional) Would you be interested in being contacted to participate in an interview or discussion group at a later date to discuss your answers to this survey in more detail?

- I consent to being contacted for an interview or discussion group. (1)
- I do not consent to being contacted. (You may still continue the survey!) (2)

Skip To: End of Block If (optional) Would you be interested in being contacted to participate in an interview or discussion... = I do not consent to being contacted. (You may still continue the survey!)

3 Thank you! Please provide a valid email below from which you can be contacted. This information will ONLY be used to contact you about interviews and discussion groups, and will not be distributed.

End of Block: Permissions

Start of Block: Demographics and Background

1 Gender

- Male (1)
- Female (2)
- Non-binary / third gender (3)
- Prefer not to say (4)
-

2 Age

- Under 18 (1)
 - 18 - 24 (2)
 - 25 - 34 (3)
 - 35 - 44 (4)
 - 45 - 54 (5)
 - 55 - 64 (6)
 - 65 - 74 (7)
 - 75 - 84 (8)
 - 85 or older (9)
-

3 Education level

- Less than high school (1)
 - High school graduate (2)
 - Some college (3)
 - 2 year degree (4)
 - 4 year degree (5)
 - Professional degree (6)
 - Doctorate (7)
-

6 What language was spoken at home when you were a child?

Q186 What language(s) did you practice using Duolingo, Mondly, or HelloTalk?

Q7 Are you currently using an in-person translator or translation software to help you with this survey?

- Yes (1)
- No (2)

Q8 How would you currently rate your proficiency/ability in the language you practiced using Duolingo, Mondly, or HelloTalk?

- Beginner (You can use some simple expressions, introduce yourself, and communicate in very simple conversations.) (1)
- Elementary (You can effectively communicate in simple and short conversations about every day things such as shopping, transportation, and other routine every-day tasks.) (2)
- Intermediate (You can hold short conversations about simple topics, handle most basic communication when travelling in a country that speaks the language, and can read and write short, simple texts.) (3)
- Upper Intermediate (You can quickly communicate in the language for most every-day tasks and simple conversations without any difficulty, you can understand the main idea of complex texts and audio segments but not many details, and you can give semi-detailed opinions on most topics.) (4)
- Advanced (You can express your ideas on most topics fluently and in a detailed manner without needing to search for words, you can understand most written texts and complex audio

segments on a wide range of topics in detail, and you can write detailed texts about subjects you are familiar with.) (5)

Proficient (You can understand almost all spoken and written language, even if technical and academic, you can express ideas quickly and fluently in conversation without strain, and you can write and speak in great detail on academic or technical topics of interest to you.) (6)

End of Block: Demographics and Background

Start of Block: Duolingo

1 Do you have any experience with the application Duolingo?

Yes (1)

No (2)

Skip To: End of Block If Do you have any experience with the application Duolingo? = No

2 Do you still use Duolingo?

Yes (1)

No (2)

3 How often did you, or do you, use Duolingo?

- Daily (1)
 - 4-6 times a week (2)
 - 2-3 times a week (3)
 - Once a week (4)
 - Rarely (5)
-

4 For how long did you use Duolingo? (Or have you been using)

- A few days (1)
 - 1-2 weeks (2)
 - Several weeks (3)
 - 2-6 months (4)
 - 6 months to a year (5)
 - More than a year (6)
-

5 How important were the following aspects of language learning to you in terms of using Duolingo?

	Extremely important (1)	Very important (2)	Moderately important (3)	Slightly important (4)	Not at all important (5)
Improve writing (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Improve reading (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Improve listening (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
improve translation skills (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
improve grammar skills (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
improve vocabulary (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Supplementing in-person classes (7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Supplementing conversation practice with language exchange or tutor. (8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Carry Forward Selected Choices from "How important were the following aspects of language learning to you in terms of using Duolingo? "



6 Overall, how would you rate the effectiveness of Duolingo in helping you to achieve your goals and expectations?

	Extremely effective (1)	Very effective (2)	Moderately effective (3)	Slightly effective (4)	Not effective at all (5)
Improve writing (x1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Improve reading (x2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Improve listening (x3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
improve translation skills (x4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
improve grammar skills (x5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
improve vocabulary (x6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Supplementing in-person classes (x7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Supplementing conversation practice with language exchange or tutor. (x8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

7 Did you take Duolingo's placement test?

- Yes (1)
- No (2)

Skip To: 12 If Did you take Duolingo's placement test? = No

8 How accurate did you feel the placement test was in determining your level?

- Extremely accurate (1)
 - Moderately accurate (2)
 - Slightly accurate (3)
 - Neither accurate or inaccurate (4)
 - Slightly inaccurate (5)
 - Moderately inaccurate (6)
 - Extremely inaccurate (7)
-

9 Did you try to skip ahead of the level that Duolingo recommended after taking the placement test?

- Yes (1)
 - No (2)
-

10 Did you decide to study at a lower level than the placement test recommended?

- yes (1)
 - No (2)
-

11 If the answer to either of the previous questions was yes, can you explain why you decided to skip ahead or start at a lower level?

12 For the level that you were assigned or decided to start at, how appropriate was the language given to you in the audio and written examples in the following categories?

(Extremely appropriate = a little challenging but still possible to do. Extremely inappropriate = either so difficult it was frustrating or so easy that you did not learn anything)

	Extremely appropriate (1)	Moderately appropriate (2)	Slightly appropriate (3)	Neither appropriate nor inappropriate (4)	Slightly inappropriate (5)	Moderately inappropriate (6)	Extremely inappropriate (7)
Grammar (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vocabulary (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Written Language (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Spoken language/audio (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Overall input (Examples of language, no what you created or produce.) (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

13 When you began to use this language software how appropriate were the exercises in terms of the language you were expected to produce in the following categories?

(Extremely appropriate = a little challenging but still possible to do. Extremely inappropriate = either so difficult it was frustrating or so easy that you did not learn anything)

	Extremely appropriate (1)	Moderately appropriate (2)	Slightly appropriate (3)	Neither appropriate nor inappropriate (4)	Slightly inappropriate (5)	Moderately inappropriate (6)	Extremely inappropriate (7)
Grammar (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vocabulary (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Written language (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Spoken language/audio (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Overall output (the language that you created in these exercises.) (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

14 Overall, how effective were Duolingo's grammar lessons and explanations in helping you to learn grammar?

- Extremely effective (1)
- Very effective (2)
- Moderately effective (3)
- Slightly effective (4)
- Not effective at all (5)

15 How well do you feel Duolingo improved your grammar for use in the following contexts?

	Extremely well (1)	Very well (2)	Moderately well (3)	Slightly well (4)	Not well at all (5)
Speaking: casual conversation (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Speaking: formal conversation (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Speaking: speeches (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Writing: emails and notes (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Writing: formal essays and articles (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

16 If you did not like the grammar instruction and thought it was not helpful, why?

If you thought that the program was extremely helpful and assisted with your grammar in contexts outside of what is listed above, why?

17 How effective would you rate Duolingo at teaching you vocabulary?

- Extremely effective (1)
- Very effective (2)
- Moderately effective (3)
- Slightly effective (4)
- Not effective at all (5)

18 How would you rate the range of vocabulary that Duolingo taught you from broad to narrow?

Broad: You feel that you learned a lot of different vocabulary words that you can utilize in a wide range of contexts and situations.

Narrow: You feel that the amount of vocabulary that Duolingo taught you is limited and specific to a small range of contexts.

- Extremely broad (1)
- Somewhat broad (2)
- Neither broad nor narrow (3)
- Somewhat narrow (4)
- Extremely narrow (5)

19 How effective would you rate the vocabulary that Duolingo taught you for specific scenarios, such as going to restaurants or travelling?

- Extremely effective (1)
- Somewhat effective (2)
- Neither effective nor ineffective (3)
- Somewhat ineffective (4)
- Extremely ineffective (5)
-

20 If you felt that the vocabulary you learned from Duolingo was not very useful or extremely useful, can you explain why?

21 How helpful do you feel Duolingo was for helping you to notice your own mistakes in the following categories?

	Extremely helpful (1)	Somewhat helpful (2)	Not helpful at all (3)
Grammar (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vocabulary (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pronunciation (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

22 How effective do you feel Duolingo was at specifically showing you your mistakes when you got a question or exercise wrong?

- Extremely effective (1)
 - Very effective (2)
 - Moderately effective (3)
 - Slightly effective (4)
 - Not effective at all (5)
-

23 What did you like or dislike about how Duolingo showed you your mistakes or corrected you?

24 When using Duolingo, how easy did you feel it was to remember new words that were introduced in the exercises?

- Extremely easy (1)
 - Moderately easy (2)
 - Slightly easy (3)
 - Neither easy nor difficult (4)
 - Slightly difficult (5)
 - Moderately difficult (6)
 - Extremely difficult (7)
-

25 How well did you feel that the materials covered in each lesson prepared you for upcoming lessons?

- Extremely well (1)
 - Somewhat well (2)
 - Not well or poorly (3)
 - Somewhat poorly (4)
 - Extremely poorly (5)
-

26 Did you feel that new lessons adequately reviewed previously covered materials to reinforce what you learned?

- yes (1)
 - Somewhat (2)
 - No (3)
-

27 How important were the following factors in your motivation to learn the language when using Duolingo?

	Extremely important (1)	Very important (2)	Moderately important (3)	Slightly important (4)	Not at all important (5)
Work (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
School (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Personal interest in English or languages in general (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Communicating with friends or family (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Being able to participate in online interaction or gaming (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Travel (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

28 After the first week, how motivated did you feel to continue to use Duolingo to practice?

- Extremely motivated (1)
- Somewhat motivated (2)
- Not very motivated or unmotivated (3)
- Somewhat unmotivated (4)
- Extremely unmotivated (5)

29 Were you able to stay motivated to use Duolingo for longer than 3 months?

- Yes (1)
 - Somewhat (2)
 - No (3)
-

30 Could you explain what motivated you to keep using Duolingo. Or, If you were unable to stay motivated, why?

31 How useful did you find the exercises and activities in Duolingo for every-day activities?

- Extremely useful (1)
 - Moderately useful (2)
 - Slightly useful (3)
 - Neither useful nor useless (4)
 - Slightly useless (5)
 - Moderately useless (6)
 - Extremely useless (7)
-

32 With the exercises and examples Duolingo provided, did you feel confident differentiating formal language from informal language in the language you practiced?

- Yes, extremely confident (1)
- Yes, somewhat confident (2)
- No, not confident at all. (3)

End of Block: Duolingo

Start of Block: HelloTalk

1 Do you have any experience using the application HelloTalk?

- Yes (1)
- No (2)

Skip To: End of Block If Do you have any experience using the application HelloTalk? = No

2 Do you still use HelloTalk?

- Yes (1)
 - No (2)
-

3 How often did you, or do you, use HelloTalk?

- Daily (1)
 - 4-6 times a week (2)
 - 2-3 times a week (3)
 - Once a week (4)
 - Rarely (5)
-

4 For how long did you use HelloTalk? (Or have you been using)

- A few days (1)
 - 1-2 weeks (2)
 - Several weeks (3)
 - 2-6 months (4)
 - 6 months to a year (5)
 - More than a year (6)
-

5 How important were the following aspects of language learning to you in terms of using HelloTalk?

	Extremely important (1)	Very important (2)	Moderately important (3)	Slightly important (4)	Not at all important (5)
Improve writing (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Improve reading (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Improve listening (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
improve translation skills (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
improve grammar skills (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
improve vocabulary (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Supplementing in-person classes (7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Supplementing conversation practice with language exchange or tutor. (8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Carry Forward Selected Choices from "How important were the following aspects of language learning to you in terms of using HelloTalk? "



6 Overall, how would you rate the effectiveness of HelloTalk in helping you to achieve your goals and expectations?

	Extremely effective (1)	Very effective (2)	Moderately effective (3)	Slightly effective (4)	Not effective at all (5)
Improve writing (x1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Improve reading (x2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Improve listening (x3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
improve translation skills (x4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
improve grammar skills (x5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
improve vocabulary (x6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Supplementing in-person classes (x7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Supplementing conversation practice with language exchange or tutor. (x8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

7 For your level, how appropriate was the language given to you by your language exchanges in audio and written examples in the following categories?

(Extremely appropriate = a little challenging but still possible to do. Extremely inappropriate = either so difficult it was frustrating or so easy that you did not learn anything)

	Extremely appropriate (1)	Moderately appropriate (2)	Slightly appropriate (3)	Neither appropriate nor inappropriate (4)	Slightly inappropriate (5)	Moderately inappropriate (6)	Extremely inappropriate (7)
Grammar (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vocabulary (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Written Language (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Spoken language/audio (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Overall input (Examples of language, no what you created or produce.) (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

8 For your level, how appropriate were the interactions with your language exchanges in what you were expected to produce in the following categories?

(Extremely appropriate = a little challenging but still possible to do. Extremely inappropriate = either so difficult it was frustrating or so easy that you did not learn anything)

	Extremely appropriate (1)	Moderately appropriate (2)	Slightly appropriate (3)	Neither appropriate nor inappropriate (4)	Slightly inappropriate (5)	Moderately inappropriate (6)	Extremely inappropriate (7)
Grammar (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vocabulary (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Written language (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Spoken language/audio (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Overall output (the language that you created in these exercises.) (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

9 How effective do you feel your language exchanges were at teaching you grammar when you used HelloTalk?

- Extremely effective (1)
- Very effective (2)
- Moderately effective (3)
- Slightly effective (4)
- Not effective at all (5)

10 How well do you feel using HelloTalk improved your grammar for use in the following contexts?

	Extremely well (1)	Very well (2)	Moderately well (3)	Slightly well (4)	Not well at all (5)
Speaking: casual conversation (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Speaking: formal conversation (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Speaking: speeches (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Writing: emails and notes (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Writing: formal essays and articles (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

11 If you did not like the grammar advice and thought it was not helpful, why?

If you thought that the program was extremely helpful and assisted with your grammar in contexts outside of what is listed above, why?

12 How effective would you rate using HelloTalk at helping you learn vocabulary?

- Extremely effective (1)
- Very effective (2)
- Moderately effective (3)
- Slightly effective (4)
- Not effective at all (5)

13 How would you rate the range of vocabulary that you learned using HelloTalk from broad to narrow?

Broad: You feel that you learned a lot of different vocabulary words that you can utilize in a wide range of contexts and situations.

Narrow: You feel that the amount of vocabulary that you learned is limited and specific to a small range of contexts.

- Extremely broad (1)
- Somewhat broad (2)
- Neither broad nor narrow (3)
- Somewhat narrow (4)
- Extremely narrow (5)

14 How effective would you rate the vocabulary that your language exchanges in HelloTalk taught you for specific scenarios, such as going to restaurants or travelling?

- Extremely effective (1)
- Somewhat effective (2)
- Neither effective or ineffective (3)
- Somewhat ineffective (4)
- Extremely ineffective (5)
-

15 If you felt that the vocabulary you learned from using HelloTalk was not very useful or extremely useful, can you explain why?

16 How helpful did you feel HelloTalk was for helping you to notice your own mistakes in the following categories?

	Extremely helpful (1)	Somewhat helpful (2)	Not helpful at all (3)
Grammar (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vocabulary (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pronunciation (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

17 How effective do you feel your language exchanges in HelloTalk were at specifically showing you your mistakes when you did something incorrectly or struggled to communicate something?

- Extremely effective (1)
 - Very effective (2)
 - Moderately effective (3)
 - Slightly effective (4)
 - Not effective at all (5)
-

18 What did you like or dislike about how HelloTalk and your exchanges in the app showed you your mistakes or corrected your language?

19 When using HelloTalk, how easy did you feel it was to remember new words that you saw in your conversations?

- Extremely easy (1)
 - Moderately easy (2)
 - Slightly easy (3)
 - Neither easy nor difficult (4)
 - Slightly difficult (5)
 - Moderately difficult (6)
 - Extremely difficult (7)
-

20 How important were the following factors to your motivation to learn the language when using HelloTalk?

	Extremely important (1)	Very important (2)	Moderately important (3)	Slightly important (4)	Not at all important (5)
Work (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
School (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Personal interest in English or languages in general (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Communicating with friends or family (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Being able to participate in online interaction or gaming (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Travel (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

21 After the first week, how motivated did you feel to continue to use HelloTalk?

- Extremely motivated (1)
 - Somewhat motivated (2)
 - Not very motivated or unmotivated (3)
 - Somewhat unmotivated (4)
 - Extremely unmotivated (5)
-

22 Were you able to stay motivated to use HelloTalk for longer than 3 months?

- Yes (1)
 - Somewhat (2)
 - No (3)
-

23 Could you explain what motivated you to keep using HelloTalk. Or, If you were unable to stay motivated, why?

24 With what you learned from your interactions in HelloTalk, did you feel confident differentiating formal language from informal language?

- Yes, extremely confident (1)
- Yes, somewhat confident (2)
- No, not confident at all. (3)

End of Block: HelloTalk

APPENDIX B: DUOLINGO QUESTION 16 RESPONSES

16 - If you did not like the grammar instruction and thought it was not helpful, why? If you thought that the program was extremely helpful and assisted with your grammar in contexts outside of what is listed above, why?

- Inefficient for several main reasons: Grammar is chopped in too small pieces, in too disconnected lessons, and it is too slowly approached (this worsened with the new "professional" course. The old one years ago was better), the pace is extremely uneven. It is far too superficial, especially the later lessons make you just memorise a few examples, rather than really practice using the grammar features. The exercises have been changed, there used to be more of the translation typing ones (=the useful ones), but now there are mainly the worthless ones, and they are the majority even at the highest difficulty level. It's now a toy, designed to be inefficient, to keep people coming and seeing ads for much longer. A standard coursebook is a much more efficient choice.
- I find Duolingo only useful for people who begin learning a language with no previous knowledge. It doesn't teach complex grammar at all, only the basics.
- The grammar was over-generalized and didn't go into depth. I remember when it taught the difference between "ser" and "estar" it said the difference was permanent versus temporary which just misled me and confused me even more.
- When I was using it, there was little to no explanation for grammar sometimes. This may have since changed.
- It gives you grammar piecemeal, without the larger language context
- didn't feel in-depth enough. Korean grammar is extremely confusing so in-depth explanations are needed to understand how to properly utilize them in context.

APPENDIX C: HELLOTALK QUESTION 11 RESPONSES

11 - If you did not like the grammar advice and thought it was not helpful, why? If you thought that the program was extremely helpful and assisted with your grammar in contexts outside of what is listed above, why?

- The other users didn't provide much helpful correction, but I don't blame them.
- Because most people didn't reply when they realized I wouldn't be teaching english or they would just want to talk in spanish, so I didn't learn a thing.
- This was all done by native speakers and so the examples they provided were extremely contextually appropriate and up-to-date. If there ever was an issue with a grammar point they mentioned it immediately and provided another example to further explain.
- The grammar advices is great. I like 'please correct me' feature

APPENDIX D: DUOLINGO QUESTION 23 RESPONSES

23 - What did you like or dislike about how Duolingo showed you your mistakes?

- The pronunciation aspect of duolingo is abysmal. Sometimes I would just say random words and it would accept it.
- Duolingo sometimes accepts mistakes as typos, not showing what your mistake was, just saying that there was a typo. Also, the way they show mistakes seems to be by underlining the first difference from the preferred answer, so what is underlined may not even be the thing that is actually wrong
- Sometimes corrections were given in a different form than what I chose to answer. Eg/ if I have to translate a sentence with 'you', and I get it wrong in the 'tu' form, it will show me the correct 'vous' form so it can't highlight what I did wrong.
- it's not always accurate. Also, sometimes it accepts little mistakes because it thinks they're typos, and then genuine typos get you a wrong answer.
- Most mistakes were just mistyped words
- There are bugs and errors. Situations like when the noun doesn't let clear if it is masculine or feminine (teacher, for example) and duolingo should accept both as answer but it doesn't.
- I did not like that some were mistakes just because they did not follow the pattern they expected, but the responses were still good. Also, on pronunciation exercises you get red highlighting but have no idea what you did wrong. You don't even have the option of listening to your recording.
- I sometimes felt like the corrections were not good at all so I decided to give it a try on Spanish, and I was true, the corrections sometimes are kind of inexact or really picky.
- Duolingo sometimes doesn't take into account more "natural" sounding sentences and can be very rigid in what it wants from translations. However, this may be because the courses are, essentially, community created.
- my mistakes were mostly typos from auto-correct; I did not learn anything about what I did wrong in the pronunciation exercises
- I started learning Korean on Duolingo as soon as it became available. It was frustrating as many grammatically correct sentences would not be accepted as they did not exactly match the expected answer

APPENDIX E: HELLOTALK QUESTION 18 POSITIVE RESPONSES

18 - What did you like or dislike about how HelloTalk and your exchanges in the app showed you your mistakes or corrected your language?

- I liked how the app allowed for multiple commenters to correct or add in examples to help assist you with correcting your mistakes.
- I liked that I could have native speakers correct my responses directly and tell me what sounds the most natural.
- I like how people try correct my post. I don't like how people interaction individual. It shows that at the end it's more like a social app other than a learning app.

APPENDIX F: HELLOTALK QUESTION 18 NEGATIVE RESPONSES

18 - What did you like or dislike about how HelloTalk and your exchanges in the app showed you your mistakes or corrected your language?

- They didn't.
- I liked how the person was kind, but their explanation was either nonexistent or not useful.
- It is just like a personal text message so that people do not usually correct partners' mistakes unless it is totally broken.
- You can talk to people but that's about it

APPENDIX G: HELLOTALK QUESTION 23 POSITIVE RESPONSES

23 - Could you explain what motivated you to keep using HelloTalk. Or, If you were unable to stay motivated, why?

- Interacting with real people, receiving active and frequent feedback etc.
- I made good friends, and i wanted to keep talking with them. That motivates me as well.

APPENDIX H: HELLOTALK QUESTION 23 NEGATIVE RESPONSES

23 - Could you explain what motivated you to keep using HelloTalk. Or, If you were unable to stay motivated, why?

- Didn't really help me learn
- It hard to get language partner who can learning language together regularly. And I don't like the feeling of glued to my phone.
- It was boring.
- I find a more direct way to speak to native speakers.

- My Russian level is so low that it is a burden to use HelloTalk.

APPENDIX I: SARAH QUOTE

Interviewer: “And then, also related to the difficulty thing, do you feel that the writing and the speaking exercises were within your skill level?”

Sarah: “No, no. Out of my skill level there.”