Comparing Duolingo and Babbel as Supplements to Classroom Foreign Language Instruction

Jacob Singleton

ISLT 9450

Dr. Jenny Bossaller

6 December 2022

Duolingo and Babbel as Supplements to Classroom Foreign Language Instruction

Background and Significance of the Research

Despite its status as a one of the oldest subjects of study, language learning has enjoyed myriad methodological changes in its history. With the advent of fast-paced and wide-spread advances in technology in past decades, these methods have changed considerably, particularly in the last ten years. Mobile- and computer-based applications, such Duolingo and Babbel, now vie for the attention of millions of daily users seeking to learn a new language, even without previous or concurrent classroom instruction in that target language.

Likewise, millions more also use these applications in addition to their own formal, in-class instruction, especially students at high schools and universities. In response, both Duolingo and Babbel advertise themselves as effective supplements to these forms of traditional learning, albeit in different ways. Babbel promotes their ability to offer differentiated additional practice, as well as tailored review of concepts in foreign languages. Duolingo makes similar claims, but has also created an online learning portal, called Duolingo Classroom, specifically for teachers to assign activities that fit their current learning objectives. While there is research pointing to the efficacy of these applications for traditional students of foreign languages, few studies feature comparisons of these applications. This study seeks to do precisely this, while also identifying the specific strengths of these applications in terms of linguistic skill acquisition and development.

Problem Statement

Both Duolingo and Babbel exist to promote and facilitate acquisition of a new language, while leveraging enjoyable gamified learning experiences and methods grounded in proven research, such as spaced repetition. In fact, much research supports the remarkable abilities of

these applications to aid in learning a new language. Despite this, almost no comparative study

of these applications and their strengths exists to guide educators in selecting one of them to use

as supplemental practice for their students, particularly those at the high school level.

Purpose of the Study

This study addresses the gap in research by comparing the uses of Duolingo and Babbel

as replacements to traditional homework in high school foreign language courses. Specifically,

this quantitative study compares the pre- and post-test performances of three separate groups of

high school students in intermediate-level German courses. One group will receive traditional

homework assignments, while the other two will each receive assignments through either

Duolingo or Babbel. This study hypothesizes that the groups using Duolingo and Babbel will

demonstrate higher proficiency in the target language, though proficiency in specific linguistic

skills will differ by application.

Literature Review

Given the vast history and numerous advancements in the methods of language learning,

this literature review is, in part, a response to this: an exploration of extant scholarship to

elucidate the utilities and effectiveness of mobile- and computer-based applications both stand-

alone and in conjunction with traditional classroom instruction, as well as the benefits,

drawbacks, and perceptions of user experience with these. The literature that follows is,

foremost, the result of database searches primarily using ERIC, PsychINFO, and Google Scholar.

The initial results were somewhat sparse, so the majority of the literature stems from

bibliographical inquiry in the articles from the database findings. Of these, many were too broad

or outdated for consideration in this study. Therefore, this review focuses on the most recent

scholarship that affords clearer understandings of the current state of the field and what research exists.

This first section of this literature review surveys the current scholarship surrounding mobile- and computer-based applications and their measured efficacy. The second section focuses more specifically on which skills see the most growth through use of these applications. The third section addresses scholarship on students' perceptions of these applications and their experience as users, as well as their attitudes and motivations. Lastly, this review will identify gaps in the scholarship and suggestions for future research.

MALL/CALL Applications and their Efficacy

There is an abundance of research on the uses and effectiveness of both Mobile Assisted Language Learning (MALL) and Computer Assisted Language Learning (CALL) – that is, the use of applications and programs, both on mobile devices (i.e. Duolingo and Babbel) and computers (i.e. Rosetta Stone), to learn a new language or support learning thereof. This is undoubtedly the most saturated field in this review and much research in the following sections is grounded originally in these initial studies of efficacy. The studies nearly unanimous support the effectiveness of CALL and MALL applications in their findings, but through different methods and lenses. Bachore (2015) surveyed a large collection of MALL research studies and claims that much of the benefit is derived from "portability" and "connectivity" (p.51). The portability of these applications allows for near unrestricted access and new opportunities to learn, whereas the connectivity allows for "various inputs from different devices" that facilitates cooperative learning (Bachore, 2015, p. 51). Other researchers claim psychological reasoning for the effectiveness. Chukharev-Hudilainen and Klepikova (2016) rooted their study in cognitive

psychology, implementing an application called Linguatorium that utilized an "adaptive tutoring algorithm" that used individualized spaced repetition to facilitate the memorization of new vocabulary in short- and long-term memory (p. 339); they found a threefold improvement in vocabulary retention with just three minutes of daily practice (pp. 349-350). Many popular MALL applications utilize similar activities rooted in these cognitive theories. Lastly, Gafni et al. (2017) suggest that the most critical factors that enhanced the learning process were "the ease of use, the gamification of the application and the ubiquity of the device and application" (p. 314).

In addition to their positive findings and evaluations, many studies also mention some drawbacks. Bachore (2015) argues that the uses of MALL applications are spatially locked, with many users unable to learn due to lack of network connection. Other studies highlight the unnatural feel and sound of applications, such as Duolingo, as well as the heavy reliance on translation (Nushi & Eqbali, 2017), which stands against most best practices in second language acquisition when used alone or as a primary learning method. It is worth noting that the majority of the scholarship handles these drawbacks as an aside and not necessarily acting in detriment to the applications.

Also important here are the settings in which the researchers conducted the studies. While some studies surveyed MALL and CALL applications in total or part isolation from classroom instruction (Grimley et al., 2011), most of them focus on MALL and CALL use as a supplement to in-class instruction. This likely stems from the stance that these applications are most effective as a supplement to in-class foreign language instruction. Though millions of active daily users utilize these applications formal instruction in the hopes of learning a new language, this is not the most effective usage. Bachore (2015) warns of the importance of guidance and explanation

of best practices for users, claiming that multi-tasking can unknowingly negatively affect their language learning (p. 52). Nushi and Eqbali (2017) take this critique a step further, suggesting that "technology does affect the language learning process, [but] it does not provide the silver bullet for that process" (p. 97). Therefore, an educator functions as a vital party in effectively implementing MALL and CALL applications for language learning.

Skill Growth and Development with MALL and CALL Applications

There are many studies that quantitatively prove the ability of mobile and computer applications to increase core language skills. However, in 2015, Bachore claimed that current research on MALL and CALL use was too focused on vocabulary, listening, and speaking with an insufficient focus on grammar and writing (p. 51). Examining the scholarship prior to this point, this appears true, though a few more recent studies have begun to fill this gap. Nonetheless, the most prevalent skill with demonstrated growth is vocabulary. Berns et al. (2015) saw a significant improvement in vocabulary retention when students used a language trainer application that their team developed in-house, while Chukharev-Hudilainen and Klepikova (2016) also reported significant improvement with spaced repetition of critical vocabulary. Similarly, Munday (2016) also explored the value of Duolingo's spaced repetition in strengthening not only student vocabulary, but also grammatical understanding. Lastly, Castañeda and Cho (2016) offer one of the few studies that illuminates the effectiveness of these applications in grammar development solely; using pre- and post-tests, they measured a marked growth in both student accuracy and confidence in conjugating Spanish verbs.

Considering the current scholarship, it is clear that vocabulary and grammar, as well as receptive skills, see considerable growth through the use of mobile and computer applications for

language learning. While Bachore (2015) remarked that speaking enjoyed sufficient coverage in the scholarship, there are noticeably fewer studies that measure productive skills (i.e. speaking and writing) and their development. One more recent study of the application, Babbel, does suggest some measured growth in speaking, though greater gains in vocabulary and grammar (Loewen et al., 2020, p. 229). Notably, this study does demonstrate a helpful and welcome shift toward broader focus in its handling of both receptive and productive tasks. Though, studies focusing on the development of written proficiency are far sparser.

Motivations, Attitudes, and Gamification

While the research strongly indicates the benefits and associated skill developments of supplemental MALL and CALL application use for language learning, the experience itself – and student perception thereof – deserve attention here. Most studies on student experience occur in tandem with research into other factors, such as skill growth, and suggest that much of this development lends to positive user experiences and attitudes towards these applications. In one of the few cross-cultural studies on this topic, Viberg and Grönlund (2013) identified a shared positive view of mobile technology for foreign language learning among Chinese and Swedish university students; the students expressed that their positive perception was grounded in three factors: individualization, collaboration, and authenticity. Munday (2016) also handles the positive attitudes based in the individualization of Duolingo in her study of A1- and B2-level university Spanish students, calling the application "flexible" and promoting of "self-directed learning" (p. 97), though she notes that more study is needed here; in support of this, her student surveys even demonstrate that over half of A1-level students completed more than the required assignments in the application (p. 94). Munday's (2016) students also displayed a strong preference for Duolingo homework over traditional homework assignments for the A1-level

students (p. 93). Meanwhile, the positive effect of collaboration on student attitudes and experiences is clear in much of the scholarship with many studies basing this in constructivist theory (Grimley et al., 2011; Nushi & Eqbali, 2017).

Gamification, however, stands out among the other positive factors for attitude and motivation. Munday's (2016) students reported that their preference for Duolingo over traditional homework was due to the fact that it made them "feel like [they were] playing a game rather than doing homework" (p. 94). Many studies have identified gamification as a positive contributor to the success and effectiveness of MALL and CALL applications. Grimley et al. (2011) used an experience sampling method to capture "real-time experience and feelings of flow" (p. 45) and found heightened feelings of flow were linked to "perceptions of a more active and challenging learning experience (p. 52), which in turn led to higher motivation. They also observed "higher anxiety states" while gaming due to the perception of challenge, though they claimed this is beneficial to learning when maintained at a low-level (Grimley et al., 2011, p. 52). Furthermore, gamification and gamified elements help to reduce the often-monotonous feeling of activities based in spaced repetition, allowing students to enjoy the activities more as they "persist in trying to get more points or rewards" (Gafni et al., 2017, p. 314).

Conclusion and Implications

Just in the past decade, numerous studies have handled the topic of mobile- and computer-based applications for foreign language learning and this trend is likely to continue as newer technologies continuously open new doors. Likewise, the foci of these studies are also wide-ranging. There is strong agreement in the scholarship that MALL and CALL applications stand to significantly enhance foreign language instruction, especially when used as a

supplement to classroom instruction. Particularly notable are the improvements in vocabulary, grammar, and receptive skills. Additionally, there is strong improvement in student motivations and attitudes, though these were dependent, in part, on the application being used. These studies also investigate several applications, though newer research has begun to shift towards new terminology with m-learning as a successor to MALL.

While this wide-cast approach in the scholarship affords a useful panorama of the field, the studies within each sub-topic are too shallow and there is much more to explore. Most of the studies are smaller scale and comparative studies are nearly non-existent. Future research should remedy this by comparing languages, cultures, levels of study and the efficacy of applications. In terms of skills, vocabulary is widely overrepresented, with some newer studies beginning to include other receptive skills (listening and reading) and grammar to an extent. Studies focusing on the growth of writing and speaking skills would be welcome here. Moreover, while some studies handle motivation, it is generally short-term and not much focuses on motivation to continue study in that language.

Lastly, Duolingo and Babbel are two of the top language learning applications and there are a wide range of articles exploring their general efficacy. There are, however, fewer studies that investigate how each of these benefit specific languages and skills and even fewer examine these applications as a supplement to in-class instruction. This study seeks to discover exactly this: how these applications support foreign language acquisition when used as supplements to in-class learning and, by means of comparison, which applications are best-suited for each skill.

Methodology

Method and Rationale

This study will employ a quantitative approach with a quasi-experimental design featuring non-equivalent groups. This design fits, given that there will be three separate participant groups and a need to measure the effects of two different language learning applications. One group will receive regular homework assignments from their high school foreign language course, while two groups will receive alternate homework assignments through popular language learning applications – one using Duolingo and the other Babbel. This study seeks to compare student pre- and post-test scores to ascertain which skills (i.e. speaking, writing, reading, listening, vocabulary, and grammar) saw the most improvement, if any, from the supplemental use of these applications. The pre-test will occur at the start of the academic semester and the post-test will occur in the final month of that same semester.

By collecting and comparing student test results, the exact effects of these applications on linguistic skill acquisition become clearer. More specifically, this study will compare student averages per skills on these tests, in particular the growth of these skills from pre- to post-test per application.

Research Question

In the scholarship many studies prove the apparent benefits of both Duolingo and Babbel, in addition to other applications, in linguistic skill acquisition but not comparatively and with a lack of representation of high school students. As a result, this study will answer the following questions:

1. When used as a replacement for traditional homework assignments, do Duolingo and Babbel facilitate comparable or improved linguistic skill acquisition in the target language for high school foreign language learners?

2. Which skills see the most and least benefit from these applications?

3. In comparison to one another, which application is best suited for each type of linguistic development (i.e. productive and receptive) in high school foreign language courses?

Variables

There are a couple of key variables at work in this study. At the center of this research, this study seeks to measure student performance per linguistic skills. Student assessment performance acts as the dependent variable in this research. Each of the two experimental groups will use a different language learning application instead of their usual homework assignments, with one utilizing Duolingo, the other Babbel. The independent variable is the type of homework assignments the students receive.

Population

This study is designed for implementation at a high school and will compare three separate classes of 20-25 students who are learning the same language at the same level. This study is not designed for high-intermediate or advanced level classes, such as those with Advanced Placement credit, as these courses are at a college-level and do not best represent the average developing language learner in a high school. Furthermore, these groups are non-equivalent and based on already existing classes at the high school. In an effort to reduce some

confounding variables, such as student motivation and failure to commit to application use, this study will measure students in Honors-level language courses in their third or fourth year.

As this study includes participants who are exclusively minors at a public educational institution, several permissions are needed. Foremost, parent permission and administrative permission at the school are required before any part of the study may commence. Parents must agree to the content and conditions of the study, while the administrative team, including the principals and any other key stakeholders at the school, need also to consent to the scope and logistical aspects of the study.

Data Collection

This study relies primarily on pre- and post-test data to measure student linguistic skill growth. Students in each participant group will take an identical pre-test to serve as a baseline for the key skills: speaking, writing, reading, listening, vocabulary and grammar. The pre-test has three sections. The first section measures receptive skills (listening and reading), as well as vocabulary and grammar, and consists of 25 questions for each of these four skills. The second and third sections measure the productive skills, testing writing and speaking, respectively. The writing and speaking sections will each include two prompts, one interpersonal and one presentational. Furthermore, this pre-test will include grammatical structures and vocabulary from the units of instruction found in the semester of learning that the study takes place in, while the receptive and productive skills will also be related to the themes in those same units, but open-ended enough to allow for spoken and written responses without in-depth knowledge of those same themes. At the close of the study period, students will take an identical post-test in order to measure the exact growth per linguistic skill per class.

Ethical Considerations

There are a few potential ethical concerns that may arise, depending on circumstances. Foremost, there is the possibility that one of the applications fails to adequately support or enhance the students' language learning. Both applications feature the ability to focus on certain topics for purposeful and directed practice. Duolingo allows educators to assign fitting assignments to students, while Babbel allows students to select the appropriate theme as a supplement to their in-class learning. Though unlikely, the classroom educator may notice a negative influence from one of these applications in relation to the study group receiving typical homework assignments and wish to discontinue or make changes to better support their students. Likewise, an educator may see significant improvement in one study group using an application and wish to offer this same experience to other students, even prior to the post-test. Students or parents may also feel they are missing out on using one of these applications and express a will to begin using one prior to the end of the study. Additionally, there is an obligation to protect the records of the students and ensure anonymity, as all participants will be minors and their educational records must be kept confidential.

Data Analysis Plan

This study will employ a one-way analysis of variance (ANOVA) test, since there are three separate groups in this comparison. A one-way ANOVA facilitates comparison of student average scores in each linguistic skill, specifically the mean score. It tests the following null and alternate hypotheses: student scores are the same with different forms of homework assignments (H0) or student scores are different with different forms of homework assignments (H1). As an omnibus test, the results will identify a significant statistical difference in one or two of the three

groups. If two groups are similar, post-hoc testing will further measure for significant statistical differences.

While the one-way ANOVA will test numerical values, not all sections of the pre- and post-test will naturally produce these. The first section of these tests will produce a numerical value, however, the second and third sections will not at first. Here, student samples will be evaluated with speaking and writing rubrics based in proficiency guidelines from the American Council for Teaching Foreign Languages (ACTFL). A numerical value will be assigned to each stage of the proficiency rubric from novice-low to advanced-high.

Timeline

This study will last one academic semester, roughly four months. The classroom teacher will already know about the study in the previous semester and have time to plan for proper implementation of the study protocols and procedures. More specifically, the teacher will prepare the necessary instructional and assessment changes to be in line with the study during this preparatory phase. During this time, permissions from parents and administration will also be obtained. The teacher will then administer the pre-test in the first instructional week of the semester and the post-test in the last week.

Conclusion

There are too few studies that research the effects of language learning applications on students enrolled in foreign language courses at the high school-level. Given the millions of students spending time acquiring another language in these formative years, there is much yet to research here. This study aims to remedy this by comparing two of the most popular language learning programs when used as a substitute for regular homework assignments. Prior research

lays the groundwork for this study in several ways. Studies indicate not only that students often

prefer these applications to regular homework assignments, but also that they are highly effective

as supplements to in-class language learning. Yet, while different studies have explored the

various strengths of these applications, there is a lack of research on which application best

supports which skills at the high school level – a critical time for many language learners. This is

the precise focus of this study.

References

Bachore, M. M. (2015). Language learning through mobile technologies: An opportunity for language learners and teachers. *Journal of Education and Practice*, *6*(31), 50–53.

Berns, A., Palomo-Duarte, M., Dodero, J. M., Ruiz-Ladrón, J. M., & Márquez, A. C. (2015). Mobile apps to support and assess foreign language learning. *Research-Publishing.Net*, 26.

Castañeda, D. A., & Cho, M.-H. (2016). Use of a game-like application on a mobile device to improve accuracy in conjugating Spanish verbs. *Computer Assisted Language Learning*, *29*(7), 1195–1204.

Chukharev-Hudilainen, E., & Klepikova, T. A. (2016). The effectiveness of computer-based spaced repetition in foreign language vocabulary instruction: A double-blind study. *CALICO Journal*, *33*(3), 334–354.

Gafni, R., Achituv, D. B., & Rachmani, G. J. (2017). Learning foreign languages using mobile applications. J*ournal of Information Technology Education: Research*, *16, 301–317.*

Grimley, M., Green, R., Nilsen, T., Thompson, D., & Tomes, R. (2011). Using computer games for instruction: The student experience. *Active Learning in Higher Education*, *12*(1), 45–56.

Loewen, S., Isbell, D. R., & Sporn, Z. (2020). The effectiveness of app-based language instruction for developing receptive linguistic knowledge and oral communicative ability. *Foreign Language Annals, 53(2), 209–233.*

Munday, P. (2016). The case for using Duolingo as part of the language classroom experience. *Revista Iberoameri- cana de Educación a Distancia, 19*, 83-101.

Nushi, M., & Eqbali, M. H. (2017). Duolingo: A mobile application to assist second language

    learning. *Teaching English with Technology, 17(1), 89–98.*

Viberg, O., & Grönlund, Å. (2013). Cross-cultural analysis of users' attitudes toward the use of

    mobile devices in second and foreign language learning in higher education: A case from

    Sweden and China. *Computers & Education*, *69*, 169–180.