Duolingo efficacy study: Beginning-level courses equivalent to four university semesters

duolingo

Xiangying Jiang[*]**, Joseph Rollinson**[*]**, Luke Plonsky**[†]**, and Bozena Pajak**[*]

Abstract

We report results of an efficacy study that evaluated the listening and reading proficiency levels of Duolingo learners who had completed the beginning-level courses in Spanish or French. The participants ($n = 225$) were learners of Spanish or French who resided in the United States, had little to no prior proficiency in the target language, and used Duolingo as their only learning tool. Their language skills were assessed using the American Council on the Teaching of Foreign Languages (ACTFL) reading and listening proficiency tests. The results indicate that Duolingo learners reached Intermediate level in reading proficiency but remained at Novice level in listening proficiency. Proficiency scores of Duolingo learners were comparable with the proficiency outcomes of university students at the end of fourth semester in language programs as reported in large-scale studies by Tschirner (2016) and Rubio and Hacking (2019), but with Duolingo learners spending only half the amount of time of four semesters of university classes. These findings demonstrate that Duolingo is effective and efficient, in addition to being accessible and free.

Keywords

Duolingo, efficacy, Spanish, French, reading proficiency, listening proficiency, foreign language

1 Introduction

Duolingo is a language-teaching platform that offers free online courses available on the web and on mobile apps. Duolingo embraces the concept of accountability and transparency, and thus aims to share the proficiency outcomes of their learners. Toward this end, the goals of the present study were to evaluate the listening and reading proficiency levels of Duolingo learners who completed the beginning-level Spanish and French courses, and to understand the amount of time users took to reach the end of the beginning-level course content. In addition, we compared their proficiency scores with semester-based proficiency outcomes of students in US-based university language programs.

The paper is structured as follows: We first provide a brief introduction to the structure of Duolingo courses and Duolingo's approach to language teaching. Then, we describe the method of data collection of the current study and report the results on learner proficiency outcomes and the time they spent learning in the app. We also provide an analysis comparing the proficiency outcomes of Duolingo users with those of university students, as reported in Tschirner (2016) and Rubio and Hacking (2019). We end the paper with a brief discussion.

1.1 Duolingo Course Structure and Approach to Language Teaching

The beginning-level content of a Duolingo course includes five sections, each of which concludes with a 'checkpoint' (see Figure 1). There are a total of 114 units in the beginning level of the Spanish course and 99 units in the beginning level of the French course, as shown in Table 1. Each unit includes 4-5 lessons at 5 difficulty levels, where higher difficulty is achieved through exercises requiring progressively more recall and production. Learners are required to complete at least one difficulty level in each row to move on to the next row.

Duolingo uses bite-sized lessons to make learning a new language feel less daunting. This approach is also supported by a growing body of research in second language acquisition and in cognitive science which demonstrates a marked advantage for 'distributed' as opposed to 'massed' practice in the target language (e.g., Carpenter, 2020; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Rogers, 2017). The incorporation of features such as streaks, crowns, gems, XPs, and leaderboards

---

[*]Duolingo, Inc.
[†]Northern Arizona University

Corresponding author:
Xiangying Jiang
Duolingo, Inc. 5900 Penn Ave
Pittsburgh, PA 15206, USA
Email: assessment@duolingo.com

Figure 1. Example Duolingo course structure.

Table 1. Number of Units in Each Section of the Duolingo Spanish and French Courses

Course section	Spanish: # of units	French # of units
1	8	10
2	26	22
3	28	21
4	25	24
5	27	22
Total	114	99

encourages Duolingo learners to return to their lessons. By doing so, Duolingo also taps into learner enjoyment, shown in a number of recent studies to be associated with increased willingness to communicate and reduced levels of anxiety (e.g., Dewaele, Witney, Saito, & Dewaele, 2018; Lee, 2020).

Duolingo takes an implicit cognitive approach as its guiding learning principle but adopts explicit learning as a supplemental method. The courses expose learners to words and grammar in sentences in the target language and allow them to gradually infer linguistic regularities from repeated exposure to and engagement with meaningful input. Meanwhile, Duolingo complements implicit learning with explicit feedback and explanations. For some concepts, explicit explanation can offer a shortcut to more efficient learning. This is especially the case for features of the target language that may be difficult to notice from input alone.

A key improvement Duolingo has made to its most popular courses is to align them with the Common European Framework of Reference (CEFR), an international standard for describing the abilities of language learners at various stages of proficiency. The CEFR guides curricular development by focusing on communicative functions, i.e., what learners actually are able to do with a language, such as asking for directions or ordering a cup of coffee.

Another important feature of Duolingo is personalized instruction. There are multiple ways that Duolingo provides individualized learning experiences. First, learners with prior knowledge of the language are encouraged to take a placement test to begin at an appropriate place in the course. Second, based on machine learning algorithms, learners receive immediate feedback when giving a wrong answer and have the opportunity to apply this feedback to a novel exercise. Third, Duolingo offers practice sessions that use spaced repetition algorithms personalized for each learner. In addition, learners are able to personalize their own learning experiences, depending on their goals and needs. Learners are allowed various routes through the course content to accommodate individual learning preferences. Some learners work to the 5th difficulty level before moving on to the next learning unit, while others only complete the first difficulty level and then choose to start the next unit. There are also learners who alternate between learning new content and reviewing previously learned content.

In addition to offering flexible and individualized learning, Duolingo is committed to offering a variety of discourse-level content, particularly for enhanced listening and reading. This longer-form, discourse-level content is valuable because it provides a real-world context for language use, demonstrates how language is organized beyond the sentence-level, and features more interactive and social aspects of the target language. For example, Duolingo Stories offers engaging narratives that scaffold learners' reading comprehension in their target language with mid-story exercises. Each story is read aloud by human voice actors so that learners also gain exposure to elements of discourse-level pronunciation, such as prosody, stress, and intonation. In addition to Stories, Duolingo offers the Duolingo Spanish and French Podcasts. These podcast episodes target high beginner and intermediate language learners. Each episode includes both target language-speaking storytellers and an English-speaking host who interjects intermittently to summarize and contextualize the target language sections.

Duolingo acknowledges that language is ultimately a tool for social interaction. Because interaction in a target language is

crucial to language learning, Duolingo is also committed to creating socially rich contexts for learners to interact in their target language. For this reason, Duolingo supports over 300 weekly language meetups around the world, each of which is hosted by members of the global Duolingo community. These Duolingo Events are opportunities for language learners to meet and practice speaking in their target language with other Duolingo users. In addition to fostering community, these events are establishing interactive spaces where key components of language learning can take place—negotiation of meaning, turn-taking, repair, etc.

Duolingo is in constant change and continuous development. New features are added daily based on large-scale data analysis and elaborate testing.

1.2   The Current Study

Millions of people use Duolingo to study a second language, but what proficiency outcomes can they expect to achieve? The current study aimed to shed light on this question by measuring the listening and reading proficiency levels of Duolingo learners who had completed the beginning-level material in the Spanish and French courses. The participants had little to no prior proficiency in the target language, and used Duolingo as their only learning tool. In particular, the current study investigated the following research questions:

1. What levels of reading and listening proficiency did Duolingo learners achieve upon reaching the end of the beginning-level Spanish and French courses? How many hours did learners spend studying on Duolingo to reach that point of their course?
2. How did Duolingo learners' reading and listening proficiency scores compare with proficiency outcomes of US-based university students in Spanish and French courses?

2   Method

2.1   Participants

The participants of the current study were 135 Spanish learners and 90 French learners on Duolingo. Below we summarize some general characteristics of the participants based on their responses to a background survey. For a detailed by-course description of participant background information, see Appendix A.

- Among 210 participants who reported age, it ranged from 18-83 with a mean of 43.99 ($SD = 15.54$);

- 74% of the participants reported speaking only English before age 6; 8% were early bilingual speakers of English and another language; and 18% of the participants did not speak English before age 6–their first languages varied widely and none of them were heritage speakers of the target language.
- 39% of the participants reported having a bachelor's degree as their highest level of education, 37% having a master's degree, and 14% having a doctoral degree;
- 78% of the participants identified themselves as Caucasian, 13% as Asian, and 3% as African American;
- 49% of the participants identified themselves as male and 48% as female.

The process and criteria of participant selection for the study were as follows:

1. They were learners with an IP Address in the United States. We excluded learners who accessed Duolingo via VPN from outside of the United States.
2. They had self-reported having no or little prior proficiency in the target language. In particular, we included learners who reported prior proficiency of 0-2 on a 0-10 scale, with 0 meaning "I have no knowledge of the language at all," and 10 indicating "I have perfect knowledge of the language." Note that Duolingo collects this information from all learners at the beginning of the course for the purposes of learner analytics, and not for course placement.
3. They received an email invitation to participate in the study the minute they completed the last session of Section 5 of their course, the Checkpoint 5 quiz. Checkpoint 5 marks the end of the beginning-level Duolingo course content.
4. They self-reported using Duolingo as the only language learning tool. They acknowledged that they did not take classes or use other programs or apps during their Duolingo course.
5. They had access to a webcam-enabled, internet-connected computer for testing. They were told explicitly that testing did not work on cell phone screens.
6. They had to be at least 18 years of age to participate.

We used data generated by learner activity in Duolingo to identify participants who met the first three criteria. Those above the age of 18 and interested in participating were asked to take a background survey. Their survey responses allowed us to narrow down our pool to only those who also met criteria 4 and 5.

2.2   ACTFL Listening and Reading Proficiency Tests

We used the ACTFL listening and reading proficiency tests as our data collection instruments. The ACTFL listening

and reading proficiency tests are standardized tests for the global assessment of reading and listening ability (ACTFL, 2013, 2014). They measure how well test-takers spontaneously comprehend the texts and discourse they read or listen to as described in the ACTFL 2012 Proficiency Guidelines. ACTFL has ten levels in its proficiency rating scale, from low to high in the order of Novice (low, mid, high), Intermediate (low, mid, high), Advanced (low, mid, high), and Superior. For the purpose of this project, we used Form E of the tests, which targets proficiency levels between Novice Low and Advanced Low. The tests were administered to each participant online by a remote proctor. The participant was asked to read or listen to 15 passages and answer three multiple-choice questions after each passage. Each test was given an ACTFL rating immediately after the test was submitted. We coded ACTFL ratings numerically by following the 1-10 point scale. See Table 2 for the mapping between the point scale and each proficiency level.

Table 2. ACTFL Ratings and Numerical Coding

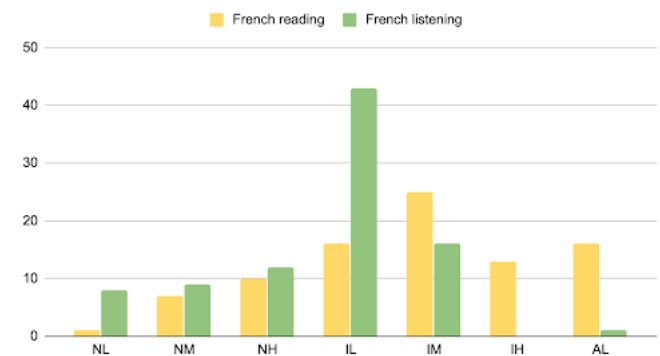| Numerical Coding	Acronym	ACTFL Rating
1	NL	Novice Low
2	NM	Novice Mid
3	NH	Novice High
4	IL	Intermediate Low
5	IM	Intermediate Mid
6	IH	Intermediate High
7	AL	Advanced Low
8	AM	Advanced Mid
9	AH	Advanced High
10	S	Superior

2.3   Data Collection Procedures

Participants completed one ACTFL proficiency test at a time, with the order of tests (reading, listening) randomized across participants. Qualified participants were first contacted by Duolingo researchers and then received an email from Language Testing International (LTI) with their test ID and instructions about how to schedule a time for the test. After the participants finished the first test, we ordered the second test for them and they were again contacted by LTI to take the second test. They went through the same process to schedule and take the test. Each participant was paid $100 after completing both tests.

Among a total of 135 Spanish-learner participants, we collected 132 reading and 131 listening scores. Among a total of 90 French-learner participants, we collected 88 reading and 89 listening scores.
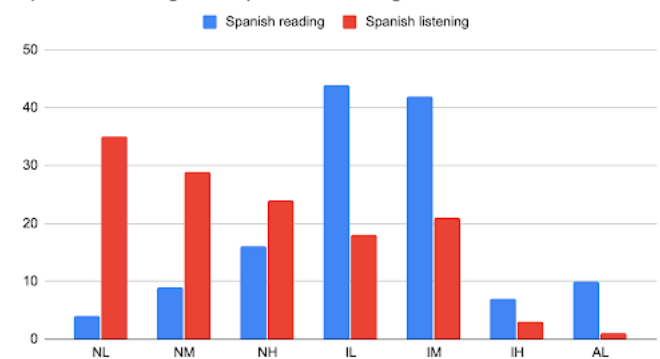
3   Results

Figure 2. Distribution of ACTFL proficiency ratings of Duolingo learners (see Table 2 for rating acronyms).

3.1   Proficiency Outcomes of Duolingo Learners

The reading and listening proficiency ratings of Duolingo learners who participated in the current study are shown in Table 3 and are visually presented in Figure 2. The ratings in Spanish reading, French reading, and French listening seemed normally distributed; however, the ratings in Spanish listening were positively skewed. Two-thirds of the Spanish listening proficiency ratings were at the Novice level.

Following the numerical coding of the proficiency ratings based on a 1-10 point scale presented in Table 2 above, we arrived at the following summary data with mean scores and standard deviations (see Table 4). Overall, Spanish and French reading scores were between Intermediate Low (4) and Intermediate Mid (5), while listening scores were at least one level behind reading. Spanish listening was approaching Novice High and French listening was at Novice High.

The reading and listening proficiency scores demonstrated the extent of effectiveness of the beginning-level Duolingo Spanish and French courses; however, another aspect of efficacy is how efficient the learning process is. In order to understand the degree of efficiency of the Duolingo Spanish and French courses,

Table 3. Distribution of Proficiency Ratings of Duolingo Learners

	NL	NM	NH	IL	IM	IH	AL	Total
Spanish reading	4	9	16	44	42	7	10	132
Spanish listening	35	29	24	18	21	3	1	131
French reading	1	7	10	16	25	13	16	88
French listening	8	9	12	43	16	0	1	89

^a [a] Columns show ACTFL Rating Acronyms (see Table 2).

Table 4. Spanish and French Reading and Listening Scores of Duolingo Learners

	N	Mean score (SD)	ACTFL rating
Spanish Reading	132	4.30 (1.34)	Intermediate Low
Spanish Listening	131	2.80 (1.54)	approaching Novice High
French Reading	88	4.82 (1.55)	approaching Intermediate Mid
French Listening	89	3.61 (1.22)	Novice High

we calculated the amount of time Duolingo learners took to reach the end of the beginning-level course content. We computed the total number of hours that the study participants spent in all Duolingo sessions in the given course, which is summarized in Table 5. The average number of hours that learners across the two courses spent studying on Duolingo was 141, and the median number was 112. French learners spent on average about 20 hours less than the Spanish learners to finish the beginning-level course, which is likely due to fewer course units in French, as reported in Table 1 above.

The numbers indicate a large variation in the amount of time learners spent learning on Duolingo. This was expected. As mentioned in the Introduction, each Duolingo course unit has five difficulty levels but learners were required to complete only one of those levels to move on to the next row. Some learners reached the 5th difficulty level in all units while others did the minimum to move along the course, thus leading to large between-participant differences in the amount of hours spent learning on Duolingo.

Based on the data presented in this section, we arrived at the following answers to the research question regarding the proficiency outcomes of Duolingo Spanish and French learners, and the time they spent studying on Duolingo:

- The Duolingo Spanish learners in the study spent a median of 125 hours working through the beginning-level course content and achieved Intermediate Low in reading and approached Novice High in listening.
- The Duolingo French learners in the study spent a median of 99 hours working through the beginning-level course content and approached Intermediate Mid in reading and achieved Novice High in listening.

3.2 Comparison with University Courses

The second goal of this study was to compare the proficiency outcomes of Duolingo learners with the outcomes of US-based university students in language courses. While the learner populations are vastly different in the two cases, we believe it informative to establish correspondences between learner proficiency outcomes across distinct educational environments.

There are two recent, large-scale studies reporting semester-based reading and listening proficiency outcomes of university courses: Tschirner (2016) and Rubio and Hacking (2019). These studies measured language proficiency of students across multiple university language programs in the United States. Both studies reported findings based on ACTFL reading and listening proficiency tests, also used by the present study, which allows for maximal comparability.

Tschirner (2016) reported listening and reading proficiency levels at different milestones of undergraduate study based on data from more than 3,000 participants learning seven languages at 21 institutions across the United States. More concretely, ACTFL listening proficiency tests and reading proficiency tests were administered to first-, second-, third-, and fourth-year students during 2014-2015. Data were collected from learners of French, German, Japanese, Italian, Portuguese, Russian, and Spanish. The main findings were reported based on listening and reading proficiency levels in Spanish and French, which made up 82% of all tests completed. In both languages, there was a steady increase in proficiency levels over the semesters in both listening and reading, but listening proficiency levels were substantially lower than reading levels. By the end of the 4th semester, on average, students reached Intermediate Low in reading proficiency (with a mean score of 4.09 in Spanish and 4.52 in French), but their listening proficiency was only approaching Novice High (with a mean score of 2.83 in Spanish and 2.65 in French).

Table 5. Hours Spent in Duolingo Sessions by Participants in the Study

Course	N	Mean	SD	Median	Min	Max	Interquartile range
Overall	225	140.56	117.01	111.99	7.27	558.63	42.52-206.12
Spanish	135	147.84	118.27	124.60	8.10	558.63	44.31-212.79
French	90	129.63	114.89	99.35	7.27	548.45	38.55-192.04

Rubio and Hacking (2019) reported findings from all three institutions of the Flagship Proficiency Initiative (University of Utah, University of Minnesota, and Michigan State University) with students enrolled in 2nd- and 4th-semester courses in Chinese, Russian, French, and Spanish between 2014 and 2016. ACTFL ratings on reading, listening, and OPIc (Oral Proficiency Interview-Computer version) were reported in the study. The results showed that after two semesters of instruction, students were consistently below the intermediate range of proficiency. After four semesters of instruction, reading reached Intermediate Low in Spanish and French (with a mean score of 4.26 in Spanish and 4.08 in French), but listening remained at Novice High (with a mean score of 3.24 in Spanish and 3.41 in French).

To compare results of the current Duolingo study with findings of these two university studies, we only focused on the reading and listening proficiency scores of the Spanish and French university students. Based on the performance of Duolingo learners reported above, we decided to make a comparison of Duolingo learners with 4th semester university students. Table 6 summarizes the scores of university students at the end of the 4th semester in the two studies. Fourth semester is the highest level in most university basic language programs and is often used as the criterion for meeting the language requirement.

To assess whether there were significant differences between university 4th-semester students and Duolingo learners, separate one-way ANOVAs on each of the four scores were carried out on the summary data of the three studies using the R statistical package (R Core Team, 2020). Only the summary data (i.e., counts, means, and standard deviations) were available to us through the published studies (Rubio & Hacking, 2019; Tschirner, 2016).

For the Spanish reading scores, there was no significant difference among the three sets of data ($F = 1.35$, $p > .05$, $\eta^2 = .003$), which suggests that Duolingo learners were not significantly different from the university students at the end of their 4th semester, as reported in Tschirner (2016) and Rubio and Hacking (2019). In Spanish listening, ANOVA showed a significant between-group difference ($F = 8.63$, $p < .05$, $\eta^2 = .020$). Pairwise comparisons showed that Duolingo learners were not significantly different from the 4th-semester university students in Tschirner (2016), but both groups scored significantly lower than the 4th-semester university students reported in Rubio and Hacking (2019).

There were significant differences among the three groups in both French reading ($F = 10.62$, $p < .05$, $\eta^2 = .037$) and listening ($F = 21.13$, $p < .05$, $\eta^2 = .072$). In French reading, pairwise comparisons showed no significant difference between Duolingo learners and the 4th-semester university students in Tschirner (2016) but both groups were significantly better than the 4th-semester university students in Rubio and Hacking (2019). In French listening, pairwise comparisons showed no significant difference between Duolingo learners and the 4th-semester university students in Rubio and Hacking (2019) but both groups performed significantly better than the 4th-semester university students in Tschirner (2016). Detailed results of pairwise comparisons can be found in Appendix B. Table 7 provides a summary of comparisons particularly relevant to the current study.

To show how proficiency scores were aligned in all three studies, we present all data combined in Figure 3.

4   Discussion and Conclusion

This study assessed the reading and listening proficiency of Duolingo learners who had completed the beginning-level material in the Spanish and French courses and compared their proficiency scores to those of 4th-semester university students on the same measures. The ACTFL proficiency scores showed that Duolingo Spanish learners reached Intermediate Low in reading and approached Novice High in listening, and Duolingo French learners approached Intermediate Mid in reading and reached Novice High in listening. These proficiency scores aligned with those of 4th-semester university students. In other words, when Duolingo Spanish and French learners reached Checkpoint 5 at the end of the beginning-level course content on Duolingo, their reading and listening proficiencies were comparable to what university students accomplished in four semesters of classes.

In addition to the evidence of effectiveness, Duolingo also offered an advantage in efficiency, including the flexibility to study a few minutes at a time, anywhere, and anytime. The median amount of time that our participants took to complete the beginning-level material was 112 hours (99 hours for French learners and 125 hours for Spanish learners), which is only half the number of hours in four semesters of university classes. There are some variations in the number of credit hours for language courses at different institutions, but typically, these

Table 6. Spanish and French Reading and Listening Scores of 4th Semester US-Based University Students

Language Skill	Study	N	Mean score (SD)
Spanish Reading	Tschirner (2016)	338	4.09 (1.64)
	Rubio & Hacking (2019)	427	4.26 (1.65)
Spanish Listening	Tschirner (2016)	317	2.83 (1.59)
	Rubio & Hacking (2019)	414	3.24 (1.37)
French Reading	Tschirner (2016)	215	4.52 (1.45)
	Rubio & Hacking (2019)	260	4.08 (1.41)
French Listening	Tschirner (2016)	203	2.65 (1.63)
	Rubio & Hacking (2019)	255	3.41 (1.33)

Table 7. Summary of Comparisons between the Current Duolingo Study and the University Studies

| Language Skill	Tschirner (2016): 4th semester	Rubio & Hacking (2019): 4th semester
Spanish Reading	No significant difference	No significant difference
Spanish Listening	No significant difference	Duolingo significantly lower
French Reading	No significant difference	Duolingo significantly higher
French Listening	Duolingo significantly higher	No significant difference
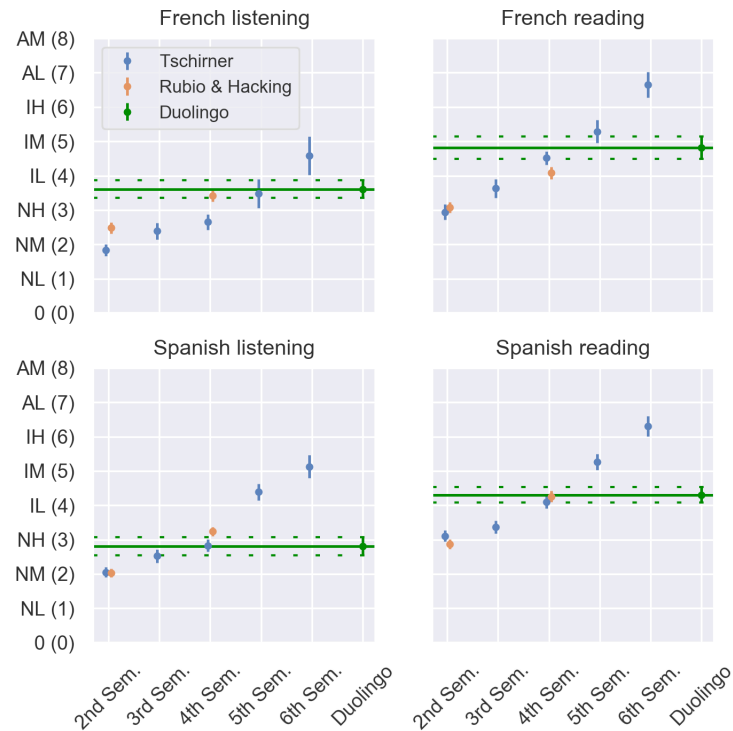
Figure 3. Comparison of ACTFL proficiency test scores in all three studies with 95% confidence intervals. The green horizontal line and confidence intervals represent performance of Duolingo learners. See Table 2 for the proficiency ratings shown on the y-axis.

language courses offer four credit hours per semester. Based on an estimate of 4 hours per week and 15 weeks per semester (not counting any hours outside of class), four semesters of university classes take 240 hours.

Some differences between our participants and the university student population, however, may have contributed to our results. The two university studies tested full-time university students from a more homogeneous age range, while our participants were more varied demographically and included

mostly post-university adults. Similarly, our participants' motivations for language learning could also be more varied than university students, who included both those studying to meet a requirement and some who would later declare majors and minors in the language.

The current study tested learners when they reached Checkpoint 5 naturalistically. For future research, treatment studies with pre- and post-test design will allow more control of participant factors that were self-reported in the present study, including prior proficiency and the exclusion of other learning tools. This study focused on listening and reading proficiency, which are both receptive skills. We did not assess learners in speaking (as in Rubio & Hacking, 2019) or writing. In the future we plan to evaluate Duolingo's effectiveness in developing learners' productive skills as well. Assessing Duolingo learners' proficiency in productive skills will allow us to determine if our learners' success in receptive skills generalize to other skills.

In sum, this study assessed the reading and listening proficiency outcomes of Duolingo learners who had little to no prior knowledge of the target language and used Duolingo as the only learning tool. The findings demonstrated that learners who finished the beginning section of the Duolingo Spanish or French course reached Intermediate level in reading proficiency and Novice High in listening proficiency. These proficiency scores of Duolingo learners were comparable with the proficiency outcomes of university students at the end of fourth semester in language programs, but with Duolingo learners spending only half the amount of time of four semesters of university classes.

Author Biographies

Xiangying Jiang holds a Ph.D. in Applied Linguistics (Northern Arizona University, 2007). She was Associate Professor of TESOL at West Virginia University before joining Duolingo in 2019. She is a senior learning scientist at Duolingo and works on learning assessment.

Joseph Rollinson is currently a staff software engineer at Duolingo, where he co-leads teams focused on learning assessment and course content development. He graduated from Carnegie Mellon University with undergraduate degrees in Computer Science and Philosophy. As an undergraduate, he performed research in intelligent tutoring systems.

Luke Plonsky is Associate Professor of Applied Linguistics at Northern Arizona University. His work, focusing primarily on second-language acquisition and research methods, has appeared in over 80 articles, book chapters, and books. Luke is Associate Editor of *Studies in Second Language Acquisition*, Managing Editor of *Foreign Language Annals*, and Co-Director of the IRIS Database.

Bozena Pajak holds a Ph.D. in Linguistics (University of California, San Diego, 2012). Before joining Duolingo in 2015, she was a Research Associate and a Lecturer in Linguistics at Northwestern University. Her research focused primarily on the acquisition of additional languages in adulthood. She is currently the Director of Learning and Curriculum at Duolingo, where she co-leads the company's Learning Area.

5 References

ACTFL. (2013). *ACTFL reading proficiency test (RPT). Familiarization manual and ACTFL proficiency guidelines 2012 – reading*. Retrieved from https://www.languagetesting.com/pub/media/wysiwyg/manuals/ACTFL_FamManual_Reading_2019.pdf

ACTFL. (2014). *ACTFL listening proficiency test (LPT). Familiarization manual and ACTFL proficiency guidelines 2012 – listening*. Retrieved from https://www.languagetesting.com/pub/media/wysiwyg/manuals/ACTFL_FamManual_Listening_2019.pdf

Carpenter, S. K. (2020). Distributed practice/spacing effect. In L.-f. Zhang (Ed.), *Oxford research encyclopedia of education*. Oxford: Oxford University Press.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354. https://doi.org/https://doi.org/10.1037/0033-2909.132.3.354

Dewaele, J.-M., Witney, J., Saito, K., & Dewaele, L. (2018). Foreign language enjoyment and anxiety: The effect of teacher and learner variables. *Language Teaching Research*, *22*(6). https://doi.org/https://doi.org/10.1177/1362168817692161

Lee, J. S. (2020). The role of grit and classroom enjoyment in EFL learners' willingness to communicate. *Journal of Multilingual and Multicultural Development*. https://doi.org/https://doi.org/10.1080/01434632.2020.1746319

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rogers, J. (2017). The spacing effect and its relevance to second language acquisition. *Applied Linguistics*, *38*(6), 906–911. https://doi.org/https://doi.org/10.1093/applin/amw052

Rubio, F., & Hacking, J. F. (2019). Proficiency vs. Performance: What do the tests show? In P. Winke & S. Gass (Eds.), *Foreign language proficiency in higher education* (pp. 137–152). Springer International Publishing AG. Retrieved from http://ebookcentral.proquest.com/lib/wvu/detail.action?docID=5622543

Tschirner, E. (2016). Listening and reading proficiency levels of college students. *Foreign Language Annals*, *49*, 201–223. https://doi.org/https://doi.org/10.1111/flan.12198

A   Appendix

Table 8. Background Information of the Participants in the Spanish and French Courses

Categories	Spanish (n=135)	French (n=90)
Age		
Mean (SD)	46.89 (18.52)	39.54 (14.96)
Prefer not to answer	8	7
Language before age 6		
English	96	70
Early Bilingual[a]	13	6
Other[b]	26	14
Highest level of education		
High school	5	7
Associate degree	5	5
Bachelor's degree	51	36
Master's degree	55	28
Doctoral degree	18	14
Prefer not to answer	1	0
Ethnicity		
African American	4	2
Asian	18	9
Causacian	105	71
Other	4	5
Prefer not to answer	4	3
Gender		
Female	76	36
Male	59	48
Other	0	4
Prefer not to answer	0	2

[a] The early bilinguals learning Spanish ($n = 13$) spoke English and one of the following languages: Chinese (4), Italian (2), Arabic, Hindi, Korean, Norweign, Polish,Taiwanese, and Turkish. The early bilinguals learning French ($n = 6$) spoke English and one of the following languages: Tagalog (2), Arabic, Spanish, Slovak, and Tamil.

[b] Spanish learners who did not speak English before age 6 ($n = 26$) spoke one of the following languages: Chinese (7), German (3), French (2), Japanese (2), Russian (2), Arabic, Dutch, Farsi, Finnish, Greek, Hindi, Tamil, Turkish, Urdu, and Zulu. French learners who did not speak English before age 6 ($n = 14$) spoke one of the following languages: Spanish (4), Russian (3), Chinese (2), Bengali, Filipino, Indonesian, Italian, and Punjabi/Hindi.

B   Appendix

Table 9. Results of Pairwise Comparisons

Language Skill	Comparison	Effect size (Hedge's g)	Confidence interval (g)
Spanish Reading	Tshirner (2016) vs. Duolingo	-.131	[-.323, .071]
	Tshirner (2016) vs. Rubio & Hacking (2019)	-.106	[-.249, .037]
	Rubio & Hacking (2019) vs. Duolingo	-.025	[-.220, .171]
Spanish Listening	Tshirner (2016) vs. Duolingo	.020	[-.184, .224]
	Tshirner (2016) vs. Rubio & Hacking (2019)	-.277	[-.423, -.130]*
	Rubio & Hacking (2019) vs. Duolingo	.297	[.0100, .494][a]
French Reading	Tshirner (2016) vs. Duolingo	-.207	[-.456, .041]
	Tshirner (2016) vs. Rubio & Hacking (2019)	.304	[.123, .485]*
	Rubio & Hacking (2019) vs. Duolingo	-.511	[-.753, -.269]*
French Listening	Tshirner (2016) vs. Duolingo	-.670	[-.929, -.420]*
	Tshirner (2016) vs. Rubio & Hacking (2019)	-.530	[-.715, -.346]*
	Rubio & Hacking (2019) vs. Duolingo	-.140	[-.318, .102]

* Denotes statistically significant difference. Hedge's g is an effect size that expresses the difference between groups in SD units. The negative sign indicates that the second group in the comparison has a higher mean. Its absolute value indicates the magnitude of the effect.