

Reading and Listening Outcomes of Learners in the Duolingo English Course for Spanish Speakers



Duolingo Research Report DRR-22-08
August 23, 2022 (10 pages)
<https://www.duolingo.com/efficacy>

Xiangying Jiang*, Bozena Pajak†

Abstract

We report results of a study that evaluated the reading and listening proficiency levels of Duolingo learners in the initial sections (CEFR A1-A2) of the English course for Spanish speakers: end of the A1 section (n = 97), middle of the A2 section (n = 94), and end of the A2 section (n = 72). The participants were learners who had little to no prior proficiency in English and used Duolingo as their only learning tool. Their language skills were assessed using the reading and listening sections of the STAMP 4S English test by Avant Assessment. The results show that learners at the end of A2 scored significantly higher than those at the end of A1, and scores increased steadily across three consecutive sections. In particular, the average scores of the learners who reached the end of A2 were Intermediate High in reading and Intermediate Mid in listening on the ACTFL scale. These findings demonstrate the effectiveness of Duolingo in developing learners' reading and listening comprehension skills.

Keywords

Duolingo, efficacy, English, reading proficiency, listening proficiency

1 Introduction

English is the most commonly studied second/foreign language around the globe, and its popularity has been increasing steadily (ICEF Monitor, 2019). Duolingo offers over 100 courses teaching 40 languages, among which English is by far the most popular. This paper reports results of a study measuring learning outcomes of learners in the Duolingo English course for Spanish speakers, thus contributing to our understanding of the efficacy of app-based education for English as a second or foreign language.

Previous work evaluated the proficiency outcomes of Duolingo's Spanish and French courses for English speakers. For example, Jiang, Rollinson, Plonsky, Gustafson, and Pajak (2021) found that after completing the Basic-level content of the Spanish or French course on Duolingo, learners achieved Intermediate Low in reading and Novice High in listening based on ACTFL Reading and Listening Proficiency Tests. The proficiency levels that Duolingo learners achieved were comparable to the levels of US-based university students after taking four semesters of Spanish or French courses. In another study, Jiang, Rollinson, Chen, et al. (2021) found that more than half of the learners met or exceeded course expectations in speaking after completing the Basic-level content of the Spanish or French course on Duolingo. However, so far no studies have examined the effectiveness of any English course on Duolingo.

The goal of the current study was to evaluate the effectiveness of the Basic-level content (CEFR A1-A2) in Duolingo's English course for Spanish speakers. The study assessed the reading and listening proficiency outcomes of learners at three different

points in the initial sections of the course. Specifically, we tested three groups of learners: (1) learners who completed the A1 course section, (2) learners who completed half of the A2 section, and (3) learners who completed the A2 section. Assessing learners at consecutive sections can give us insights into the progress learners make as they move along the course, as well as information about the quality of the curriculum.

2 Duolingo's Course Structure

Like its Spanish and French courses for English speakers, Duolingo's English course for Spanish speakers is also aligned to the *Common European Framework of Reference* (CEFR), an international language proficiency standard that defines learning goals for Basic (levels A1 and A2), Independent (B1 and B2), and Proficient (C1 and C2) users (Council of Europe, 2001). Figure 1 is a sketch of the CEFR levels.

Duolingo's Basic-level content includes five CEFR-aligned sections: introduction (pre-A1/A1.0), the first half of A1 (A1.1), the second half of A1 (A1.2), the first half of A2 (A2.1), and the second half of A2 (A2.2). Duolingo is transitioning to a new home screen design of its course structure (see Figure 2), but the

*Lead Learning Scientist, Duolingo, Inc.

†VP of Learning and Curriculum, Duolingo, Inc.

Corresponding author:

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

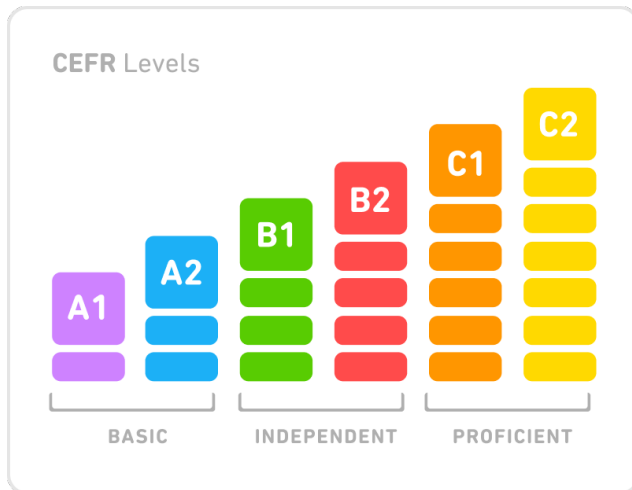


Figure 1. A sketch of the CEFR levels.

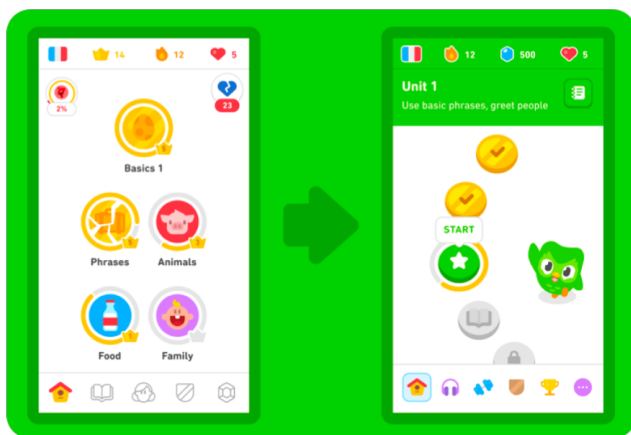


Figure 2. Duolingo's course structure home screen is transitioning to a new design. The left side shows the original tree structure and the right side shows the new path structure.

data for the current study was collected when the participants were using the original version. To help readers see how research findings based on the original version might be applicable to the new version of the course structure, a brief comparison of the two versions follows.

In the original version of the Duolingo course structure, each section of the course concludes with a “checkpoint.” Each circle in the original home screen represents a “skill,” which includes 4-5 lessons at 5 difficulty levels (“crown” 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.

In the new version of the course structure, a circle on the path is roughly equivalent to one “crown” level (difficulty level) in the original version. The content of each original “skill” is now

condensed into three levels in the new structure, and levels from different “skills” are interspersed throughout the path together with additional circles for review, as well as reading and listening practice. The content in the new path is organized into small units. See this [blog post](#) (Munson, Yu, Rajgarhia & Noh, 2022) for information about the new Duolingo learning path and the differences between the original and new designs of the home screen.

Table 1 shows the correspondence between sections, “skills,” and units in the two versions of the English course for Spanish speakers. As shown in the table, no changes were made to the scope of the CEFR-aligned course sections. The main difference is in the organization of the content. For example, it takes 116 “skills” in the tree structure but 81 units in the path structure to cover content through A2. Duolingo’s English course for Spanish speakers currently has content through B1. The shaded rows in Table 1 represent the sections that were evaluated in this study.

The lessons in the English course include several activity types targeting learning and practice in vocabulary, grammar, reading, listening, writing, and speaking. In addition to word- and sentence-based activities, Duolingo lessons include activity types that rely on comprehension of short passages or dialogues, which increase the amount of input in the target language. To facilitate listening and speaking development, Duolingo provides learners with many opportunities to listen to the target language and speak it out loud. All English course content is accompanied by audio and learners are allowed to play the audio at varied speeds as often as they need. In addition, speech recognition technology is used for all speaking exercises in order to provide learners with feedback.

Lessons are the primary method of teaching new material on Duolingo, but other modes of learning are available outside of the main course structure. For example, learners can complete generalized practice sessions, which review content they have studied throughout the entire course. For “skill”-specific practice, learners can return to any “skill” for which they have completed all difficulty levels in order to refresh their knowledge of a particular functional topic or grammar concept. Another relevant feature is Stories, which provides discourse-level reading and listening comprehension practice, reinforcing and enriching learners’ knowledge by situating the lesson content in everyday contexts. Due to the large degree of user autonomy in navigating the platform, there is considerable variation in the types of sessions that learners choose to complete. As a result, there can be substantial variation among individual learners on both the percentage of the course material they engage with, as well as on the total amount of time spent learning.

Table 1. A Comparison of the Original Tree Structure and the New Path Structure for the Duolingo English Course for Spanish Speakers

CEFR-aligned course sections	Original tree structure		New path structure	
	Number of “skills”	Cumulative number of “skills”	Number of units	Cumulative number of units
A1.0	10	10	10	10
A1.1	29	39	19	29
A1.2	22	61	16	45
A2.1	29	90	19	64
A2.2	26	116	17	81
B1.1	28	144	23	104
B1.2	28	172	18	122
B1.3	25	197	19	141
B1.4	27	224	19	160

3 Research Questions

The current study evaluated the reading and listening proficiency outcomes of learners in Duolingo’s English course for Spanish speakers when participants reached the end of Section A1.2, Section A2.1, or Section A2.2. Section A2.2 marks the end of the Basic-level (CEFR A-level) course content. In particular, we asked the following research questions:

1. How do learners’ reading and listening proficiency outcomes compare at three consecutive Basic-level course sections?
2. Do these proficiency scores provide cross-sectional evidence of continuous growth?
3. Specifically, what levels of proficiency do learners achieve at the end of A2, the Basic-level content?

4 Methods

4.1 Participants

The participants of the study were 263 Duolingo learners in the English course for Spanish speakers. To receive the initial invitation email, learners’ self-reported prior proficiency had to be 0-2 on a 0-10 scale. Duolingo asks all learners how much language they know when they first start learning on Duolingo for the purposes of learner analytics. In the 0-10 prior proficiency scale, 0 represents “I have no knowledge of the language at all,” and 10 indicates “I have perfect knowledge of the language.” In addition to limited or no prior proficiency, learners’ latest session had to be within 3 rows before or after Checkpoint 3, 4, or 5. These checkpoints correspond to the end of Section A1.2, Section A2.1, and Section A2.2. All participants were 18 years of age or older.

The qualification for participation was further restricted based on the answers to two questions in the background survey, which was included in the invitation email. These two questions asked learners whether they took classes or used other apps or programs to learn English while they were studying English on

Duolingo. Only those who reported that they did not take classes or use other apps or programs during their Duolingo course were deemed eligible to participate. See Appendix A for a summary of participant characteristics.

5 Instruments

5.1 The Background Survey

The background questionnaire included questions related to participants’ language background, reasons for learning the language, level of education, age group, and whether they took classes or used other programs/apps during the time they used Duolingo. The answers to the latter questions confirmed eligibility for participation (see Participants above). The background survey was translated into Spanish for accessibility purposes. The responses to the survey questions are summarized and shown in Appendix A.

5.2 The STAMP 4S English Test: Reading and Listening Sections

The test used in this study was a commercial standardized test called STAMP 4S provided by [Avant Assessment](#). The acronym STAMP stands for **S**tandards-Based **M**easurement of **P**roficiency, and 4S refers to the four sections/skills of reading, writing, listening, and speaking. Avant Assessment also allows the use of separate sections of the test, for example, reading and listening only. The STAMP 4S English test is an online, ACTFL-aligned, computer-adaptive test of English language proficiency accredited by the American Councils on Education (ACE). The test is scored on the STAMP scale (1-9), which is aligned with the ACTFL proficiency scale, as seen in Figure 3. The ACTFL proficiency scale ([ACTFL, 2012](#)) includes four broad levels: Novice, Intermediate, Advanced, and Superior. Each of the first three broad levels is further divided into Low, Mid, and High. The STAMP scale of 1-9 is aligned to nine ACTFL sublevels: Novice (Low, Mid, High), Intermediate (Low, Mid, High) and Advanced (Low, Mid, High).

STAMP Level 1	STAMP Level 2	STAMP Level 3	STAMP Level 4	STAMP Level 5	STAMP Level 6	STAMP Level 7	STAMP Level 8	STAMP Level 9
Novice			Intermediate			Advanced		
Low	Mid	High	Low	Mid	High	Low	Mid	High

Figure 3. The alignment of the STAMP scale with the ACTFL proficiency scale (Santos, 2022, p. 2).

The Reading and Listening sections of the STAMP 4S English test were used in this study. Each section consists of 30 multiple-choice questions, which assesses test-takers' ability to comprehend a variety of written or spoken texts used for general communicative purposes in English. The two sections of the test take 60-75 minutes to complete. They are scored automatically on a scale of 1-9 (i.e., STAMP level 1-9 or Novice-Low to Advanced-High on the ACTFL scale) and scores are available immediately after a test-taker completes these sections. Each Reading and Listening question has an associated benchmark level. Test-takers experience questions at various levels because the reading and listening sections are computer-adaptive. Appendix B shows topics and general student characteristics associated with the benchmark levels.

In addition to receiving a STAMP level score in these two sections, test-takers also receive a scaled score. Scaled scores provide a more fine-tuned view of a test-taker's proficiency. Figure 4 shows an interpretation of the scaled scores in relation to the ACTFL proficiency scale. According to Avant Assessment (Santos, 2022), the internal consistency reliability coefficients (Cronbach's alpha) of the Reading and Listening sections of the STAMP 4S English test are 0.89 and 0.90.

6 Procedures

An email soliciting participation was sent to a random sample of Duolingo learners who were at most three rows before or after Checkpoint 3, 4, or 5 in the English course for Spanish speakers, if their self-reported prior proficiency was 0-2 (see Participants section). Learners aged 18 and above interested in participating completed a background survey that allowed us to verify eligibility and collect additional demographic information. Among the survey responders, those who reported that they had taken classes or used other apps/programs to learn English during the time they used Duolingo were disqualified from participation.

Qualified participants were notified and invited to take the Reading and Listening sections of the STAMP 4S English Test paid for by Duolingo. Data were collected during seven test windows on a rolling basis, each lasting for two weeks (from initial call for participation to taking the test). Remote human proctors from Avant Assessment were present for each

Benchmark	Group	Reading	Listening
Advanced	High	617 – 660	622 – 671
	Mid	597 – 616	599 – 621
	Low	580 – 596	582 – 598
Intermediate	High	541 – 579	534 – 581
	Mid	523 – 540	521 – 533
	Low	502 – 522	501 – 520
Novice	High	461 – 501	466 – 500
	Mid	449 – 460	455 – 465
	Low	331 – 448	343 – 454

Figure 4. Interpretation of STAMP 4S English reading and listening scaled scores in relation to ACTFL proficiency scale (Avant Assessment, 2022).

scheduled testing session. Each participant received \$50 and their score report after taking the test. Table 2 shows the data collection funnel.

Table 2. Data Collection Funnel

Email sent	Survey responded	Test eligible	Test started	Test completed
32935	2034	818	299	263

7 Results

As explained in the Instruments section, the participants' reading and listening performances were evaluated in both STAMP levels and scaled scores. The STAMP levels are in an ordinal scale of 1-9, which corresponds to the ACTFL proficiency scale of Novice Low (1) to Advanced High (9). Each STAMP level corresponds to a range of scaled scores, which further

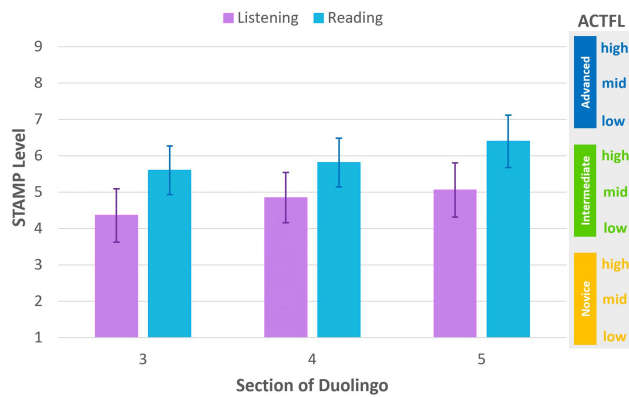


Figure 5. Reading and listening proficiency levels of participants at each course section. The confidence intervals are standard deviation units.

differentiate same-level learners and provide a more fine-tuned understanding of their proficiency. We first report the results of the study using STAMP levels, and then scaled scores, both interpreted based on Avant’s alignment with the ACTFL proficiency scale (see Figure 3 and Figure 4 above).

7.1 Reading and Listening Proficiency in STAMP Levels

Table 3 and Figure 5 show the reading and listening proficiency scores of the participants in STAMP levels. The participants who completed the course through Section 3 (end of A1) averaged Intermediate Mid in reading and Intermediate Low in listening. The participants who completed the course through Section 4 (middle of A2) also averaged Intermediate Mid in reading and Intermediate Low in listening, but their average STAMP Level scores were higher than those who completed Section 3 in both reading and listening. The participants who completed the course through Section 5 (end of A2) showed even higher scores, especially in reading. Specifically, the average reading proficiency was at Intermediate High (STAMP Level 6) and the average listening proficiency was at Intermediate Mid (STAMP Level 5). The data shows a gradual increase in both reading and listening scores across the three consecutive course sections, but the reading scores were consistently higher than the listening scores. The correlation between reading and listening scores was significant but moderate ($r = .60, p < .001$).

Linear regression models were built to test for significant differences in STAMP level scores across sections, while controlling for prior proficiency, which has a small significant correlation with reading ($r = 0.17, p < 0.01$) but not listening. Separate models were built for comparing Section 5 to Section 4 and Section 5 to Section 3, and separate models were built for reading and listening scores (4 models total). For reading, the scores of Section 5 were significantly higher than scores both at

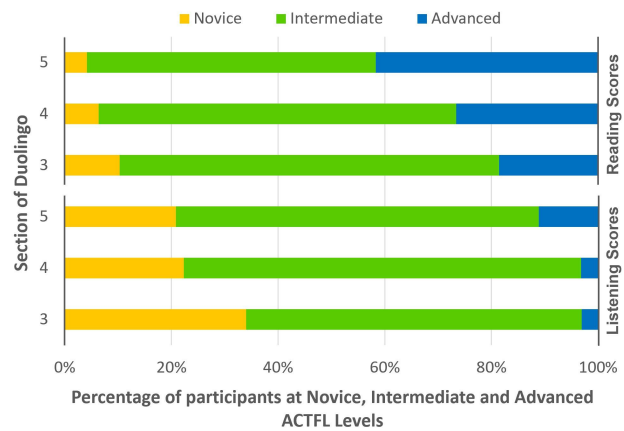


Figure 6. Percentages of participants at Novice, Intermediate, and Advanced ACTFL levels in reading and listening at the completion of each course section.

Section 4 ($\beta=0.668; SE=0.243; t=2.751; p<0.01$) and Section 3 ($\beta=0.879; SE=0.252; t=3.492; p<0.001$). For listening, learners at Section 5 have significantly higher scores than learners at Section 3 ($\beta=0.843; SE=0.240; t=3.506; p<0.001$).

Table 4 shows the number of participants who scored at each STAMP level across the scale in reading and listening. Based on the alignment between STAMP and ACTFL scales (see Figure 3 above), we combined the numbers into three categories: Novice, Intermediate, and Advanced for each course section. Figure 6 provides a visual comparison across course sections based on percentages.

Figure 6 demonstrates a pattern in the participants’ performance in both reading and listening: as the amount of completed course content grew larger, the percentage of participants scoring at Novice level decreased and the percentage at Advanced level increased. At completion of the Basic-level content (through A2.2), the percentage of participants who scored at Advanced level was 42% in reading and 11% in listening.

7.2 Reading and Listening Proficiency in Scaled Scores

Table 5 shows the average scores of the participants at each course section and their alignment with the ACTFL scale (see Figure 4). Compared to the interpretation based on STAMP level scores in Table 3, the scaled scores aligned one sublevel higher in the ACTFL scale in most cases. Because each STAMP level includes a range of scaled scores, some scores fall at the higher end of the range while others at the lower end of the range. In this case, the higher alignment probably indicated that more scaled scores fell at the higher end of their level. The summary in Table 5 shows a smaller increase in average reading scores across sections, which were all within the same range of Intermediate High. Listening scores showed a bigger increase

Table 3. Reading and Listening Proficiency of Participants in STAMP Levels across Course Sections

CEFR-aligned course sections	N	Reading		Listening	
		Mean STAMP Level (SD)	ACTFL scale	Mean STAMP Level (SD)	ACTFL scale
3 (A1.2)	97	5.60 (1.47)	Intermediate Mid	4.36 (1.34)	Intermediate Low
4 (A2.1)	94	5.81 (1.38)	Intermediate Mid	4.85 (1.34)	Intermediate Low
5 (A2.2)	72	6.43 (1.43)	Intermediate High	5.01 (1.41)	Intermediate Mid

Table 4. Number of Participants Scoring at Each STAMP Level in Reading and Listening

	CEFR-aligned course sections	N	STAMP Level								
			1	2	3	4	5	6	7	8	9
Reading	3 (A1.2)	97	0	0	10	14	12	43	9	5	4
	4 (A2.1)	94	0	1	5	13	8	42	17	6	2
	5 (A2.2)	72	0	0	3	5	2	32	15	8	7
Listening	3 (A1.2)	97	0	1	32	28	8	25	1	2	0
	4 (A2.1)	94	0	1	20	20	8	42	2	1	0
	5 (A2.2)	72	0	0	15	15	5	29	7	1	0

across sections, and they were Intermediate Low, Intermediate Mid, and Intermediate High, respectively. The increasingly better performance across sections suggests that progression in the course leads to higher proficiency levels, especially in listening. At the completion of Section 5 (A2.2), the Basic-level course content, the study participants averaged Intermediate High in both reading and listening.

8 Discussion

8.1 Summary of Findings

This study evaluated three groups of learners at the completion of three consecutive sections in the Basic-level content of Duolingo's English course for Spanish speakers. The comparisons between the participants who completed Section 3 (A1.2) and those who completed Section 5 (A2.2) showed a difference of approximately one ACTFL sublevel in both reading and listening proficiency based on scores in STAMP levels, going from Intermediate Mid to Intermediate High in reading and Intermediate Low to Intermediate Mid in listening, or two ACTFL sublevels in listening based on STAMP scaled scores, going from Intermediate Low to Intermediate High. This growth pattern was also evident in the decrease of percentage scores at the Novice level and increase at the Advanced level across the sections in both reading and listening. Specifically, the participants who completed Section 5 (A2.2), the Basic-level content, scored on average at Intermediate High in reading and Intermediate Mid in listening based on the STAMP level scores, or at Intermediate High in both reading and listening based on the scaled scores.

The findings also showed that participants' listening proficiency level was about one ACTFL sublevel lower than reading proficiency, which replicated the findings in previous studies (Jiang, Rollinson, Plonsky, et al., 2021; Rubio & Hacking, 2019; Tschirner, 2016). Although listening and reading are both receptive skills, the comprehension processes have been found to differ (Wolf, Muijselaar, Boonstra, & de Bree, 2019). Listening comprehension demands a higher level of attention, exerts a heavier load on working memory, and requires the ability for speedy decoding and processing of transient audio input (see, e.g., Bloomfield et al., 2010; Wallace, 2022). In contrast, learners' decoding process in reading is facilitated by the availability of visually presented text (Spoden, Fleischer, & Leucht, 2020; Vandergrift & Baker, 2015). As a result, listening comprehension is often more challenging than reading comprehension for second language learners, especially those at beginning levels. This study suggests that there are still opportunities for improvement in developing learners' listening comprehension skills on Duolingo.

Although the data for this study was collected based on the original version of the Duolingo course structure, we expect the findings to straightforwardly translate to the new version of the Duolingo path structure (see Figure 2). Given that Duolingo's course content is aligned to the CEFR, the curriculum for each section of the course is not changing. In other words, the material covered in each section is consistent across the two versions (see Table 1). The original version offers five difficulty levels for each "skill" but learners are only required to complete the lowest difficulty level. In a similar study by Jiang, Rollinson, Plonsky, et al. (2021), the majority of the lessons completed by the study participants were at the lowest level of difficulty and

Table 5. Reading and Listening Proficiency of Participants in STAMP Scaled Scores across Course Sections

CEFR-aligned course sections	N	Reading		Listening	
		Mean Scaled Score (SD)	ACTFL scale	Mean Scaled Score (SD)	ACTFL scale
3 (A1.2)	97	549.41 (38.20)	Intermediate High	517.37 (29.04)	Intermediate Low
4 (A2.1)	94	553.68 (33.24)	Intermediate High	527.97 (31.45)	Intermediate Mid
5 (A2.2)	72	568.54 (36.32)	Intermediate High	543.51 (35.85)	Intermediate High

some learners rarely completed higher levels. The new version of the Duolingo path structure condenses each original “skill” into three levels and puts each level directly in the learner’s study path, together with practice skills and Stories. Most of the learners will go over more content review for each “skill” in the new version of the path structure, which makes us believe they will learn at least as much, if not more. However, future studies should be conducted with participants in the new path structure to confirm this statement.

8.2 Limitations and Future Directions

This study has a few limitations. First, the inferences about proficiency development across sections were based on analysis of cross-sectional instead of longitudinal data. In this study, different groups of learners were tested when they reached the end of their respective sections. In the future, longitudinal data testing the same learners at the completion of multiple sections would allow us to better control for participant-related factors.

Although the design of the study presents some level of ecological validity because the participants reached the end of these sections naturally and independently, future research might benefit from more controlled designs such as a pre- and post-test design or a comparison-group design. These designs will allow more control of learning time and participant factors that were self-reported in the present study.

Furthermore, the skills of reading and listening assessed in the study are both receptive. Learners were not assessed in productive skills such as speaking (as in Jiang, Rollinson, Chen, et al. (2021) for Duolingo learners of Spanish and French) and writing, or overall proficiency. Future studies should evaluate Duolingo’s effectiveness in developing English learners’ productive skills or overall proficiency. Doing so will lead to a better understanding of whether and to what extent Duolingo English learners’ success in receptive skills can also be observed in productive skills or overall proficiency.

9 Conclusion

In sum, this study evaluated the reading and listening proficiency outcomes of Duolingo English learners who had little to no prior knowledge of the target language and used Duolingo as the only learning tool. Three groups of learners from Duolingo’s English

course for Spanish speakers were assessed at the end of Sections A1.2, A2.1, and A2.2, respectively. The findings demonstrated that participants who completed more course sections were more likely to achieve higher reading and listening scores than those who completed fewer sections. The group that finished the Basic-level content (A2.2) reached, on average, Intermediate High in reading and Intermediate Mid in listening proficiency based on Avant Assessment’s STAMP levels, or Intermediate High in both reading and listening based on STAMP scaled scores. These proficiency scores indicate that the Duolingo English course for Spanish speakers is effective in developing learners’ reading and listening skills. These findings, although based on the original version of Duolingo’s course structure, are expected to be applicable to the new version of Duolingo’s path structure.

Author Biographies

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

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 VP of Learning and Curriculum at Duolingo.

Acknowledgements

We would like to thank Erin Gustafson, Audrey Kittredge, and Lucy Skidmore for their help in collecting and analyzing the data, as well as providing feedback on the report.

10 References

- ACTFL. (2012). *ACTFL Proficiency Guidelines*. Retrieved from <https://www.actfl.org/sites/default/files/guidelines/ACTFLProficiencyGuidelines2012.pdf>
- Bloomfield, A., Wayland, S., Rhoades, E., Blodgett, A., Linck, J., & Ross, S. (2010). *What makes listening difficult: Factors affecting second language listening comprehension [technical report]*. Center for Advanced Study of Language (CASL), University of Maryland.
- Council of Europe. (2001). *Common European Framework of References for Languages: Learning, teaching, assessment*. Cambridge University Press, New York, NY. Retrieved from <https://rm.coe.int/1680459f97>
- ICEF Monitor. (2019). *The world's changing language landscape*. Retrieved from <https://monitor.icef.com/2019/10/the-worlds-changing-language-landscape/>
- Jiang, X., Rollinson, J., Chen, H., Reuveni, B., Gustafson, E., Plonsky, L., & Pajak, B. (2021). *How well does Duolingo teach speaking skills? [White paper]*. Retrieved from <https://duolingo-papers.s3.amazonaws.com/reports/duolingo-speaking-whitepaper.pdf>
- Jiang, X., Rollinson, J., Plonsky, L., Gustafson, E., & Pajak, B. and. (2021). Evaluating the reading and listening outcomes of beginning-level Duolingo courses. *Foreign Language Annals*, 54, 974–1002. <https://doi.org/10.1111/flan.12600>
- 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.
- Santos, V. (2022). *The STAMP 4S English Test Reading and Listening Sections*. [Technical report shared by Avant Assessment].
- Spoden, C., Fleischer, J., & Leucht, M. (2020). Converging development of English as foreign language listening and reading comprehension skills in German upper secondary schools. *Frontiers in Psychology*, 11, 1116. <https://doi.org/10.3389/fpsyg.2020.01116>
- Tschirmer, E. (2016). Listening and reading proficiency levels of college students. *Foreign Language Annals*, 49, 201–223. <https://doi.org/10.1111/flan.12198>
- Vandergrift, L., & Baker, S. (2015). Learner variables in second language listening comprehension: An exploratory path analysis. *Language Learning*, 65, 390–416. <https://doi.org/10.1111/lang.12105>
- Wallace, M. (2022). Individual differences in second language listening: Examining the role of knowledge, metacognitive awareness, memory, and attention. *Language Learning*, 72, 5–44. <https://doi.org/10.1111/lang.12424>
- Wolf, M., Muijselaar, M., Boonstra, A., & de Bree, E. (2019). The relationship between reading and listening comprehension: Shared and modality-specific components. *Reading and Writing*, 32, 1747–1767. <https://doi.org/10.1007/s11145-018-9924-8>

A Appendix

Table 6. Characteristics of the Study Participants

Characteristics	% of Participants (N = 263)
Age	
18-34 years	34.60%
35-54 years	49.05%
55-74 years	16.35%
Home language before age 6	
Only Spanish	94.30%
Only one language, but not the target language assessed in the study	3.42%
More than one language, but not the target language assessed in the study	1.52%
Other	0.76% ¹
Highest level of education	
Bachelor's degree	40.68%
Master's degree	12.17%
Doctoral degree	2.66%
Other	44.49%
Primary reason for learning the language	
For fun / leisure	29.66%
For travel	42.59%
For memory / brain acuteness	43.73%
For job-related purposes	54.75%
For social purposes	42.97%
For school	49.81%
Other	30.04%

1. Two participants answered "English" as their home language before age 6, which indicates that these participants are heritage speakers. We included them in the analysis because they met all of our participant selection criteria.

B Appendix

Benchmark Level	Topics	Characteristics
NOVICE	<ul style="list-style-type: none"> • Self • Calendar/Time • Colors/Shapes • Pets/Animals • School/Classroom • Weather/Seasons • Clothing • Food/Beverage • Family/Friends • Home • Places/Geography • Community • Daily Routines • Shopping/Stores • Leisure/Activities 	<p>Students who are reading or listening at Novice proficiency are characterized by:</p> <ul style="list-style-type: none"> • Reliance of learned phrases and basic vocabulary • Ability to recognize the purpose of basic texts • Can understand a core of simple, formulaic utterances
INTERMEDIATE	<ul style="list-style-type: none"> • Health • Holidays/Celebrations • Occupations/Professions • Transportation Travel/Vacations • Future plans • Culture • Contemporary Issues <ul style="list-style-type: none"> • Current events • Economics • Literature • Science • Social Science • History • <i>Plus more in-depth aspects of Novice topics</i> 	<p>Students who are reading or listening at Intermediate proficiency are characterized by:</p> <ul style="list-style-type: none"> • In reading, ability to understand the main ideas and explicit detail in everyday language • Ability to use language knowledge to understand information in everyday materials • Can follow short conversations and announcements on common topics and answer questions about the main idea and explicitly stated details
ADVANCED	<ul style="list-style-type: none"> • Arts • Politics • Religion • Math • <i>Plus more in-depth aspects of Novice and Intermediate topics</i> 	<p>Students who are reading or listening at Advanced proficiency are characterized by:</p> <ul style="list-style-type: none"> • Can understand and use language for straightforward informational purposes • Can understand the content of most factual, non-specialized materials intended for a general audience • Can understand the content of most spoken factual, non-specialized language

Figure 7. Topics and test-taker characteristics associated with Benchmark levels.