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

Xiangying Jiang, Duolingo, United States
Elise Hopman, Duolingo, United States
Ryan Peters, Duolingo, United States

WorldCALL 2023 – CALL in Critical Times
Conference Proceedings

Abstract
This study evaluated the reading and listening proficiency levels of 81 English language learners who completed the Basic content (CEFR A2) of the Duolingo English course for Japanese speakers. Participants self-reported having little to no prior proficiency in English and using Duolingo as their only learning tool. Their language skills were assessed with the reading and listening sections of the STAMP 4S English Test by Avant Assessment. The results show that, on average, learners at the end of A2 scored Intermediate High in both reading and listening, which is one sublevel above the expected learning outcomes (Intermediate Mid) of the course. These findings provide evidence of the effectiveness of Duolingo's English course for Japanese speakers in developing learners' reading and listening comprehension skills.

Keywords: Duolingo, Efficacy, English, Reading Proficiency, Listening Proficiency, Japanese Speakers

WorldCALL Conference 2023 in Chiang Mai, Thailand

Introduction

In Japan, English is a required school subject for students starting from third grade in elementary school. Surprisingly, after taking English classes for over 8 years at school and, for some, even more years at a university, the English proficiency of learners in Japan has been ranked at the lower end among countries based on international proficiency test results such as TOEIC, TOEFL, and IELTS (Nuttall, 2019). For example, on average performance of TOEIC test takers worldwide (Educational Testing Service, 2022a, 2022b), Japan ranked #29 among 41 countries in the reading and listening test, #22 among 26 countries in the speaking test, and #21 among 24 countries in the writing test. As English continues to be the most popular international language, there is a need for learners in Japan to gain a higher proficiency in English. This study investigated the impact of learning English via Duolingo, an app-based language learning tool, on English reading and listening proficiency of learners in Japan.

This paper reports results of a study measuring reading and listening abilities of Duolingo learners in its English course for Japanese speakers when they finish the first four sections in the course. Content in the first four sections of the course is designed to teach to the A2 level based on the *Common European Framework of Reference* (CEFR).

The CEFR is an international language proficiency standard that defines learning goals for Basic (A1-A2), Independent (B1-B2), and Proficient (C1-C2) users (see Figure 1; Council of Europe, 2001).
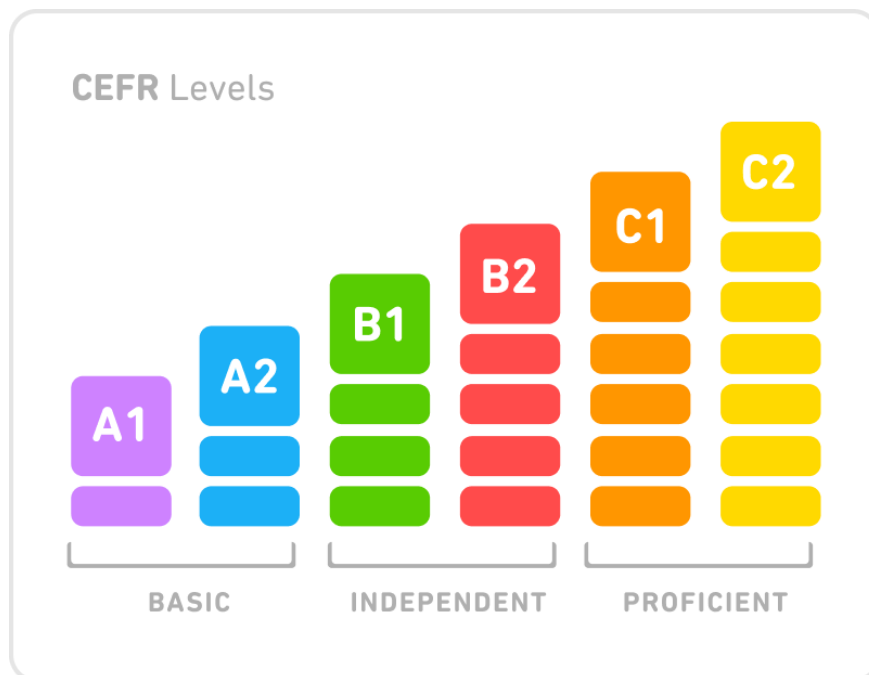

Figure 1: A Sketch of the CEFR Levels

Duolingo course structure is organized based on sections and units. There are four sections in the Basic content: a brief intro section, two sections of A1 content, and a longer section of A2 content. Each section is composed of units, and an average unit contains around 25 lessons, each lasting about 2-3 minutes. Table 1 shows the number of units in each section of the CEFR Basic content. To reach the end of A2, learners need to complete a total of 80 units in

the course. Duolingo's English course for Japanese speakers currently has content beyond A2, but, in this study, learners were tested at the end of A2.

CEFR Basic Level	Duolingo Course Section	Number of Units	Cumulative Units
Pre-A1 (Intro)	Section 1	5	5
A1	Section 2	17	22
	Section 3	18	40
A2	Section 4	40	80

Table 1: CEFR Basic-level Content in the Duolingo English Course for Japanese Speakers

The content in each unit includes lessons that either introduce new material or review previously covered content, as well as short stories. Lessons include several activity types targeting vocabulary, grammar, reading, listening, writing, and speaking. 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. Review lessons provide personalized spaced repetition of the material to ensure each learner practices their weaker areas. Finally, short stories provide discourse-level reading and listening comprehension practice, reinforcing and enriching learners' knowledge by situating the lesson content in everyday contexts.

Research Questions

The current study evaluated the reading and listening proficiency outcomes of learners in Duolingo's English course for Japanese speakers when the participants completed the first four sections of the course (Basic content through A2). The research question addressed in this study is:
- What levels of reading and listening proficiency do participants achieve when they reach the end of A2 in Duolingo's English course for Japanese speakers?

Methods

Participants

The participants of the study were 81 learners in the Duolingo English course for Japanese speakers.

To qualify for study participation, learners had to meet the following criteria:
1. Learners' self-reported English proficiency was either "I don't know any English" or "I know basic words and phrases" when they first started the English course on Duolingo. The prior proficiency scale had four levels. Only learners whose recorded prior proficiency was at Level 1 or Level 2 were included in the study. The two higher

levels were "I can understand simple conversations" and "I am intermediate or advanced." Duolingo collects self-reported prior proficiency information from all learners at the beginning of the course for the purposes of learner analytics. Except for two participants who reported knowing no English, all the participants in this study reported knowing basic words and phrases.

2. Learners reached the end of the A2 section (Section 4) in the Duolingo English course for Japanese speakers. This meant that their latest completed session in that course had to be within Units 78-80, where Unit 80 is the end of A2. Because most participants started the course with some knowledge of English, they began to learn at different places in the course after taking the placement test. The placement data showed that 37% of the participants started learning in Section 1, 30% in Section 2, and 33% in Section 3 (see Table 1 about course sections).

3. Learners were 18 years of age or older (as self-reported in the background survey included in the initial invitation email).

4. Learners did not take English classes or use other apps or programs to study English while they were learning English on Duolingo (as self-reported in the background survey).

Instruments

The Background Questionnaire

The background questionnaire included questions related to participants' language background, reasons for learning the language, highest 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 questionnaire was conducted in Japanese. The responses to the survey questions are summarized in Appendix A in English.

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*[1]. The acronym STAMP stands for **Sta**ndards-Based **M**easurement of **P**roficiency, and 4S refers to the four sections/skills of reading, writing, listening, and speaking. 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 Reading and Listening sections of the STAMP 4S English Test were used in this study. Each section consists of 30 multiple-choice questions, which assess test-takers' ability to comprehend a variety of written or spoken texts used for general communicative purposes in English. 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 (Avant Assessment, 2023a). The two sections of the test take 60-75 minutes to complete. The test is scored automatically in two ways: in ordinal ratings on a scale of 1-9 (STAMP levels) and in interval scaled scores. According to Avant Assessment

---

[1] https://avantassessment.com/stamp

(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, respectively.

Both the STAMP level ratings and scaled scores are aligned with three broad levels on the proficiency scale of the American Council on the Teaching of Foreign Languages (ACTFL): Novice, Intermediate, and Advanced, with each being further divided into Low, Mid, and High (ACTFL, 2012). As shown in Figure 2, 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

Note: From Santos (2022, p. 2). Reprinted with permission.
Figure 2: The Alignment of the STAMP Scale with ACTFL Proficiency Scale

Figure 3 shows a mapping of the scaled scores in relation to the ACTFL proficiency scale.

Compared to ordinal proficiency ratings, the scaled scores provide a more fine-tuned view of a test-taker's proficiency, thus more precise. For example, if a test-taker receives STAMP Level 6 (Intermediate High) in reading, the corresponding scaled score is within the range of 541-579. For this reason, the main findings of this study are reported based on scaled scores.

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

Note: From Avant Assessment (2023b). Reprinted with permission.

Figure 3: STAMP English Reading and Listening Scaled Scores in Relation to ACTFL Proficiency Scale

Procedures

An email soliciting participation was sent to all Duolingo learners in the course who pre-qualified for study participation (criteria 1-3 listed in the Participants section above). 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 between November, 2022 and June, 2023 during 16 test windows on a rolling basis, each lasting two weeks (from initial call for participation to taking the test). Remote human proctors from Avant Assessment were present for each scheduled testing session. Each participant received $75 and their score report after taking the test. Table 2 shows the data collection funnel.

Email sent	Survey responded	Eligible	Test started	Test completed
2734	528	254	82	81

Table 2: Data Collection Funnel

Results

As explained in the Instruments section, 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 differentiate same-level learners and provide a more precise understanding of their proficiency. Table 3 shows the number of participants who scored at each STAMP (and ACTFL) level across the scale in reading and listening. Figure 4 provides a visual presentation of score distributions in reading and listening.

STAMP	ACTFL		Reading	Listening
1		Low	0	0
2	Novice	Mid	0	0
3		High	6	6
4		Low	6	18
5	Intermediate	Mid	21	11
6		High	36	41
7		Low	4	3
8	Advanced	Mid	5	2
9		High	3	0

Table 3: Distribution of Scores at Each STAMP and ACTFL Level in Reading and Listening
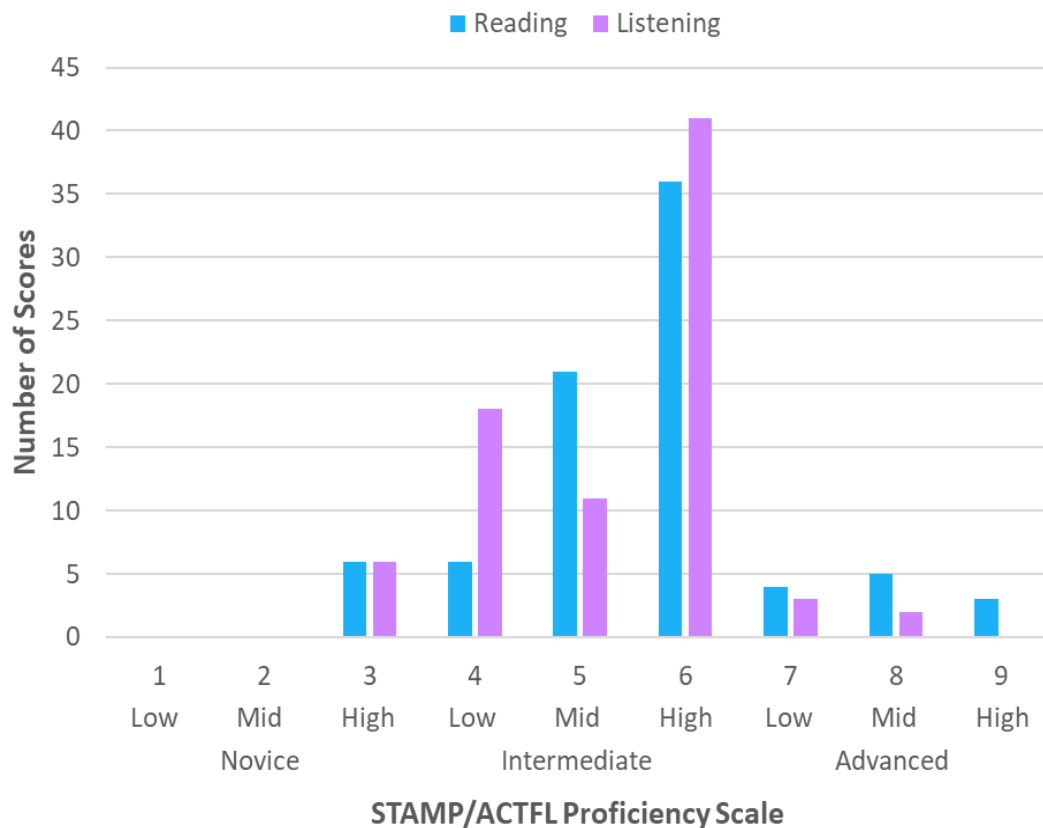

Figure 4: Reading and Listening Score Distributions by STAMP/ACTFL Level
at the End of A2

The score distributions demonstrate that the majority of the learners scored in the intermediate range, with Intermediate High (Level 6) being the mode of the distribution (i.e., the most frequent score).

Table 4 shows the average scaled scores of the participants and their alignment with the ACTFL scale. At the completion of A2 in the course, the study participants averaged 550 in reading and 539 in listening, which correspond to Intermediate High in both skills (see Figure 3).

	Mean scaled score (SD)	ACTFL scale
Reading	550 (33.78)	Intermediate High
Listening	539 (28.10)	Intermediate High

Table 4: Reading and Listening Proficiency of Japanese-Speaking Participants in STAMP Scaled Scores

Discussion

This study evaluated learners at the completion of the first four sections (Basic content through A2) of Duolingo's English course for Japanese speakers. On average, the participants scored Intermediate High in both reading and listening. The findings were consistent with those of Jiang and Pajak (2022) and Jiang et al. (2023), which reported the reading and listening proficiency of learners in Duolingo's English courses for Spanish speakers and for Portuguese speakers. The consolidated evidence boosts our confidence that Duolingo's Basic content (through A2) in its English courses is effective in developing learners' reading and listening skills.

What does a Proficiency Level of **Intermediate High** Mean?

There are two major frameworks in the field of language learning and assessment: the CEFR and the ACTFL guidelines. As well-established educational standards, they both provide the basis for curriculum development, test development, and test score interpretation. This study involved both standards. The Duolingo English courses were aligned with the CEFR and the participants were assessed when they completed the A2 content, while the interpretation of their test scores was based on the ACTFL proficiency guidelines.

The questions are how CEFR A2 relates to ACTFL Intermediate High and whether the participants met the expected course outcomes in this study. Although the two frameworks have co-existed for more than 20 years, few empirical studies investigated the correspondence between them. Currently, the interpretation of the STAMP 4S English Test scores is only aligned to the ACTFL proficiency guidelines. Avant Assessment does not provide a concordance between their STAMP scale and the CEFR. However, Avant Assessment (Santos, 5/22/2023, personal communication) expects that "a STAMP level 6 (Intermediate High) will map to a CEFR B1, given our current work on developing STAMP 4S CEFR tests and previous research in this area." Based on the work of other standardized international tests, a score of CEFR B1 is comparable to a score of 550-785 in the TOEIC

Listening and Reading Test (total 990), 42-72 in the TOEFL iBT (total 120), and 4.5-5.0 in the IELTS (total 9.0).

ACTFL (n.d.) published an empirically-based alignment between the two frameworks and uses it to assign CEFR levels to their own assessments. For the ACTFL reading and listening proficiency tests, CEFR A2 corresponds to ACTFL Intermediate Mid, and CEFR B1 is aligned to ACTFL Intermediate High (B1.1) and Advanced Low (B1.2). Based on these correspondences, the reading and listening proficiency of the participants in the current study is one ACTFL sublevel above our expected course outcomes of Intermediate Mid. In other words, the participants of this study scored above A2 and at early B1 when they completed the A2 content on Duolingo's English course for Japanese speakers, which demonstrates the effectiveness of the course in developing learners' reading and listening skills.

At the level of Intermediate High in reading and listening proficiency, what are learners able to comprehend? Avant Assessment provides some benchmark characteristics of learners based on three broad proficiency levels (see Appendix B). At the intermediate level of reading or listening, learners are assessed on topics related to health, transportation, culture, and contemporary issues such as current events, economics, literature, science, social science, and history. In reading, they are able to use language knowledge to understand the main ideas and explicit details in everyday language. In listening, learners are able to follow short conversations and announcements on common topics and answer questions about the main ideas and explicitly stated details.

Limitations and Future Directions

This study assessed learners when they reached the end of A2 naturally and independently, with no experimental manipulation, so this design presents a high level of ecological validity. However, self-reported data was used to determine participant eligibility (e.g., prior proficiency, whether or not participants were taking classes and using other tools during the use of Duolingo). Future research would benefit from more controlled designs such as a pre- and post-test design or a comparison-group design. These designs would allow more control of learning time as well as 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 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.

Conclusion

In sum, this study evaluated the reading and listening proficiency of Japanese-speaking learners of English who self-reported having little to no prior knowledge and using Duolingo as their only learning tool. The findings demonstrated that participants who completed the first four sections (CEFR A2) of the Duolingo English course for Japanese speakers scored, on average, Intermediate High in both reading and listening, which are above our course expectations. These proficiency outcomes indicate that the Duolingo English course for Japanese speakers is effective in developing learners' reading and listening skills, similarly to

what was found for the Duolingo English courses for Spanish speakers (Jiang & Pajak, 2022) and for Portuguese speakers (Jiang et al., 2023).

Appendices

Appendix A: Characteristics of the Study Participants

Characteristics	% of Participants (N=81)
Age	
18-34 years	23.46%
35-54 years	64.20%
55-74 years	12.35%
Home language before age 6	
Only Japanese	100%
Highest level of education	
Bachelor's degree	53.09%
Master's degree	13.58%
High School	16.05%
Associate degree	6.17%
Trade School	6.17%
Other	4.94%
Primary reason for learning English	
For social purposes	39.51%
For job-related purposes	39.51%
For fun/leisure	34.57%
For school	28.40%
For memory / brain acuteness	17.28%
For travel	19.75%

Appendix B: Topics and Test-taker Characteristics Associated with the Benchmark Levels

Benchmark Level	Topics	Characteristics
NOVICE	• 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	Students who are reading or listening at Novice proficiency are characterized by: • Reliance of learned phrases and basic vocabulary • Ability to recognize the purpose of basic texts • Can understand a core of simple, formulaic utterances

INTERMEDIATE	• Health
	• Holidays/Celebrations
	• Occupations/Professions
	• Transportation Travel/Vacations
	• Future plans
	• Culture
	• Contemporary Issues
	  • Current events
	  • Economics
	  • Literature
	  • Science
	  • Social Science
	  • History
	
	• *Plus more in-depth aspects of Novice topics*
		Students who are reading or listening at Intermediate proficiency are characterized by:
		
		• 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	• Arts • Politics • Religion • Math • *Plus more in-depth aspects of Novice and Intermediate topics*	Students who are reading or listening at Advanced proficiency are characterized by: • 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

Note: From Avant Assessment (2023a). Reprinted with permission.

References

ACTFL. (2012). ACTFL Proficiency Guidelines. Retrieved on July 21, 2022 at
https://www.actfl.org/resources/actfl-proficiency-guidelines-2012

ACTFL. (n.d.). Assigning CEFR ratings to ACTFL assessments. Retrieved on May 22, 2023
at
https://www.actfl.org/uploads/files/general/Assigning_CEFR_Ratings_To_ACTFL_A
ssessments.pdf

Avant Assessment. (2023a). STAMP benchmarks & rubric guide. Retrieved on August 1,
2022 from https://avantassessment.com/stamp-benchmarks-rubric-guide

Avant Assessment. (2023b). STAMP Scaled Scores Guide. Retrieved on August 1, 2022
from https://avantassessment.com/stamp-scaled-scores-guide

Council of Europe. (2001). Common European Framework of Reference for Languages:
Learning, teaching, assessment. New York, NY: Cambridge University Press.

Educational Testing Service. (2022a). ETS TOEIC L&R 2021 report on test takers
worldwide. Retrieved on May 23, 3023 from https://www.ets.org/content/dam/ets-
org/pdfs/toeic/toeic-listening-reading-report-test-takers-worldwide.pdf

Educational Testing Service. (2022b). ETS TOEIC S&W 2021 report on test takers
worldwide. Retrieved on May 23, 2023 from https://www.ets.org/pdfs/toeic/toeic-
speaking-writing-report-test-takers-worldwide.pdf

Jiang, X., & Pajak, B. (2022). Reading and listening outcomes of Duolingo English course
for Spanish speakers. [white paper]. https://duolingo-
papers.s3.amazonaws.com/reports/duolingo-english-efficacy-whitepaper.pdf

Jiang, X., Peters, R., & Pajak, B. (2023). Reading and listening outcomes of Duolingo
English course for Portuguese speakers. [white paper] https://duolingo-
papers.s3.amazonaws.com/reports/duolingo-efficacy-english-reading-listening-
whitepaper.pdf

Jiang, X., Rollinson, J., Chen, H., Reuveni, B., Gustafson, E., Plonsky, L., & Pajak, B.
(2021). How well does Duolingo teach speaking skills? [white paper]
https://duolingo-papers.s3.amazonaws.com/reports/duolingo-speaking-whitepaper.pdf

Nuttall, L. (2019). Comparative education: Why does Japan continue to struggle with
English? *The Reading Matrix: An International Online Journal, 19*(1), 74-92.

Santos, V. (June, 2022). The STAMP 4S English Test Reading and Listening Sections.
[Technical report shared by Avant Assessment].