

# Novice learners improve oral fluency with AI tutors in Kaizen Languages app

Ryo Nakayama and Yuki Yoshimura

## Introduction

Japanese language learning methods are becoming more diverse in the last couple of decades. Pew Research Center (2020) shows that 89% of college students own a smartphone in the U.S. Research at the state university in the U.S. reported that 45.2% of students who study Japanese as a foreign language use their smartphone 1-3 hours per day, 29% spend 4-7 hours per day, and 21% spend 8 hours or more per day (Yoshimura & Shiomi, 2016). While excessive internet use has been a widely known issue lately, another study confirms that college students in Malaysia use their smartphone mainly for Internet (98.2%), social media (90.9%), and applications (90.9%) (Fook, Narasuman, Aziz, Mustafa, & Han, 2021). To further look at how college students use their smartphone on the internet, a study conducted in the U.S demonstrates that the majority of college students spend their time on Communication (99.2%), Entertainment (98.4%), and Social Media (96.7%) (Penglee, Christiana, Battista, & Rosengerg, 2019). However, such excessive use of internet is not only for entertaining purposes, but also for learning. Marasuman, Aziz, Mustafa, and Han (2021) shows that 43.6% of college students spend 25% of their daily smartphone usage time to study. With the increasing number of language learning apps available, the use of such apps to study a foreign language becomes more common than before. Learners who try to learn a foreign language with the use of such apps expect to improve their speaking skills more than any other skills such as writing and reading (Fernando, 2018). While a number of foreign language learning apps are available now, not many apps focus on speaking practice. Kaizen Languages, released by AB Language in 2019, is one of the unique apps that provides an opportunity for users to practice speaking without any prior reading and writing knowledge of Japanese characters, i.e., hiragana, katakana, and kanji. Therefore the current study examines whether learners are able to improve their speaking skill particularly oral fluency in relation to the accuracy and the frequency of the app use when they have no prior reading and writing knowledge in Japanese. For this reason, the study uses the Kaizen Languages app as a tool for participants to learn Japanese only from the speaking practice.

## Kaizen Languages

Kaizen app has five functions: (1) lessons (conversation practice with AI tutors), (2) flashcards (vocabulary in each lesson), (3) writing practice for hiragana, katakana, and kanji, and (4) learning progress check. It has 61 lessons which are from beginning to advanced levels. The flashcards correspond to each lesson accordingly. When users complete a lesson, the app shows two options where they either proceed to a flashcard section or proceed to the next lesson. Writing sections provide learners with practice in Japanese hiragana, katakana and Kanji. You can set up your notification and goal each day in the Learning Progress section. While a variety of functions are available in the Kaizen app, the current research focuses on learners' speaking skill with no prior knowledge in reading and writing. Thus the study only uses the function of conversation practice with AI tutors in a dialogue format.

Dialogues programmed in each lesson starts with an AI tutor, and users are expected to respond to it in a grammatically and phonemically meaningful way. Since users have no prior reading knowledge in Japanese, sentences orally presented in a dialogue are described in romanized characters with an option to show in Japanese characters. Users are expected to interpret what they hear based on translation in English, a suggestion that provides phonetic information, and instruction that provides contextual clues. Lessons are programmed to include similar expressions so users have a chance to review similar expressions and grammatical structures in multiple lessons in a row. However, acquiring vocabulary is relatively challenging only with an orally spoken dialogue practice, hence the use of flashcards is helpful.

## Research Questions and Research Method

The current study examines (1) how the frequency of app use relates to improving fluency and accuracy in oral speech, and (2) when learners start improving their fluency, and how many repetitions are required to improve fluency. Ten adults participated in the study. They are all native English speakers with no prior experience learning Japanese. Participants first attended the research orientation session to learn how to use the Kaizen Languages app and understand the research procedures. After which, they studied Japanese with lessons 1 through 19 in the Kaizen app on their own phone for three weeks. The participants chose one of two groups and practiced according to their assigned lesson schedule. Group 1 had five lessons at a time, seven times a week. Group 2 had 11 lessons three times a week. In this study, five participants chose Group 1 and five chose Group 2. The total time to practice in a week was expected to be 70 to 75 minutes for both Group 1 and 2, and the time was set to be roughly equal. They were supposed to study each lesson five times in total. After three weeks of practice, participants took an oral speaking test. The results are analyzed based on the data from the app usage logs and the oral speaking test.

## Results and Analysis

### (1) Practice Frequency and Practice Fluency

This study looked at five components, i.e., practice frequency, practice fluency, practice accuracy, test fluency, and test accuracy. Practice fluency was measured by comparing two items. One is the participant's fastest time from the repetitive practice to complete each lesson, and the other is the control time which is the duration of time that a native Japanese speaker took to complete a lesson. The controlled time was the average time from the two attempts on the same lesson by the native speaker. The control time in every lesson is set to the value of 1.00. The participants' any exceeding time from the control time will have a value of 1.00 plus the ratio of the exceeding time. This means that the higher the value in practice fluency, the lower the actual fluency.

Practice frequency shows the total number of practice each participant completed including all lessons. When students completed all 19 lessons five times, the practice frequency should be 95. Any participants who have more than 95 times of practice frequency means that they practiced a certain lesson(s) more than five times. The figure 1 below shows the relationship between practice frequency and practice fluency.

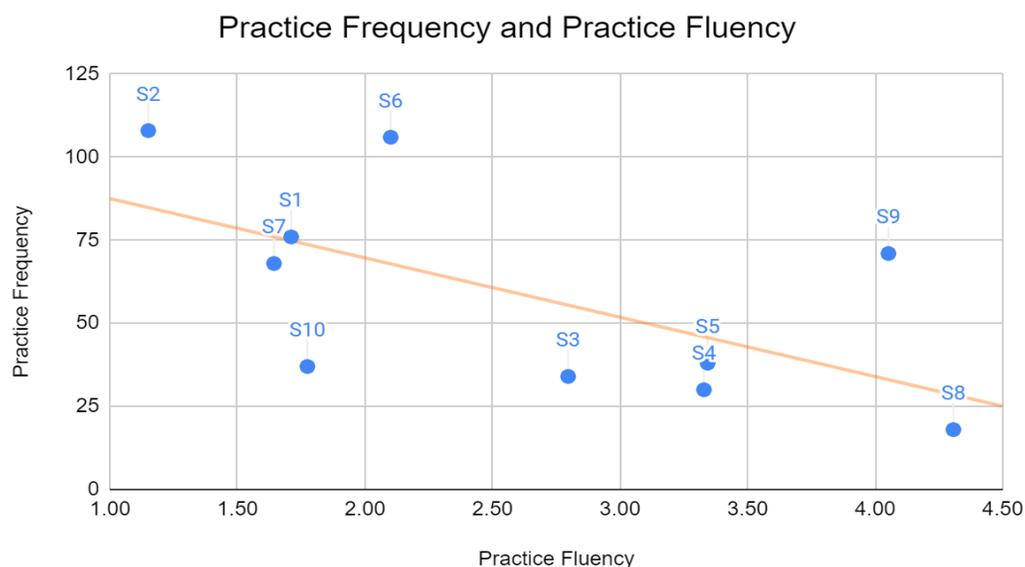


Figure 1: Practice Frequency and Practice Fluency

The smaller values in practice fluency on the x-axis indicate a higher fluency. Practice frequency on the y-axis shows the total number of practice the participants completed. For example, S2 (student 2) and S6 had the higher practice frequency whereas they also had the shorter time to complete practices compared to other participants. The shorter time here means that they were able to speak Japanese fluently in Kaizen practice. The results also show that most students who completed a larger number of practice sessions tend to be more fluent in Kaizen practice than those who completed fewer practice sessions.

The result shows that the more frequently students practice, the higher fluency they achieve. Another question that arises in relation to practice fluency is whether students who practice more show higher fluency not only in the Kaizen app practice, but in spontaneous conversation.

## (2) Practice Frequency and Test Fluency

This section looks at how practice frequency relates to fluency in the speaking test. The study measured participants' oral fluency by counting the number of morae per second in each sentence they produced. This method was also used in Kormos 2006 and Lennon 1990, in which they used morae per minute instead of morae per second. In this study, most sentences the participants produced were shorter than one minute per sentence, and thus morae per second was used. Learners took the speaking test twice, and the better performance from each participant was used for analysis. The value of morae per second shows the fluency in each sentence produced in all of the questions in the speaking test. Figure 2 shows the relationship between practice frequency and fluency in the speaking test, i.e., test fluency.

In Figure 2, points that are located further to the right indicate a higher test fluency. For example, S6, S1, S3, and S2 display high fluency in the speaking test. All four of these participants, except for S3, have more practice frequency compared to other students. The result indicates that the more students practice, the higher test fluency they achieve. On the other hand, test fluency also tends to be low when the practice frequency is low. This tendency is similar to the relationship between practice frequency and practice fluency illustrated in Figure 1.

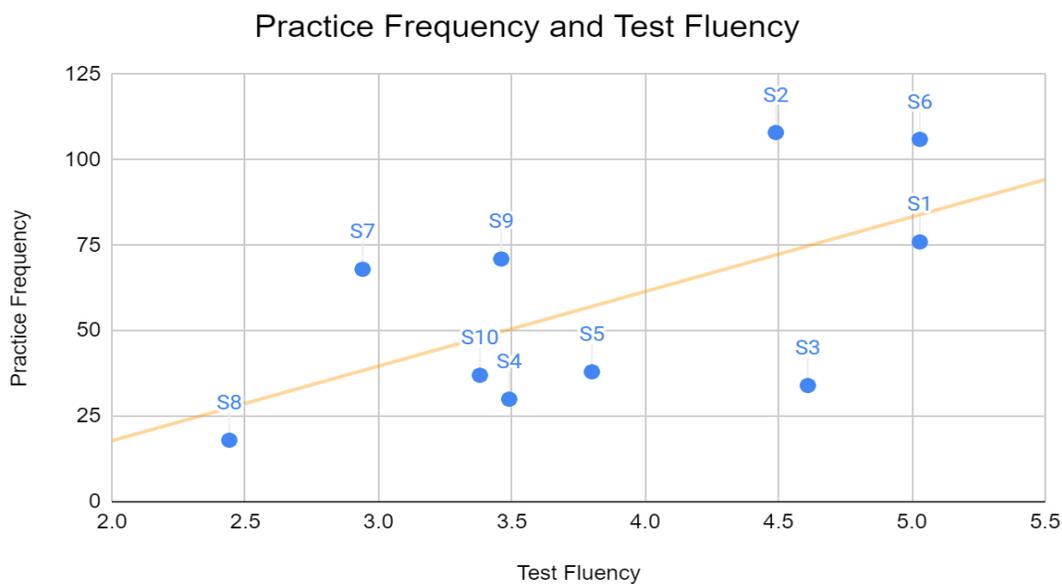


Figure 2: Practice Frequency and Test Fluency

### (3) Test Fluency and Practice Fluency

This section looks at whether students who have high fluency in the practice sessions are also highly fluent in the speaking test. Figure 3 below shows the relationship between test fluency and practice fluency. The data on the bottom right shows the highest fluency both in the practice sessions and in the speaking test. The result is rather sporadic because practice fluency is the duration of time that students take to complete a dialogue in each lesson from start to finish which may include time participants get distracted during practice.

In contrast, the data extraction method to measure test fluency follows Lennon's (1990) measurements for speech rate, i.e., unpruned speech rate. Unpruned speech rate includes all self-correction, asides, and so forth. If a student uses a large quantity of self-correction, the test fluency would be measured to be high.

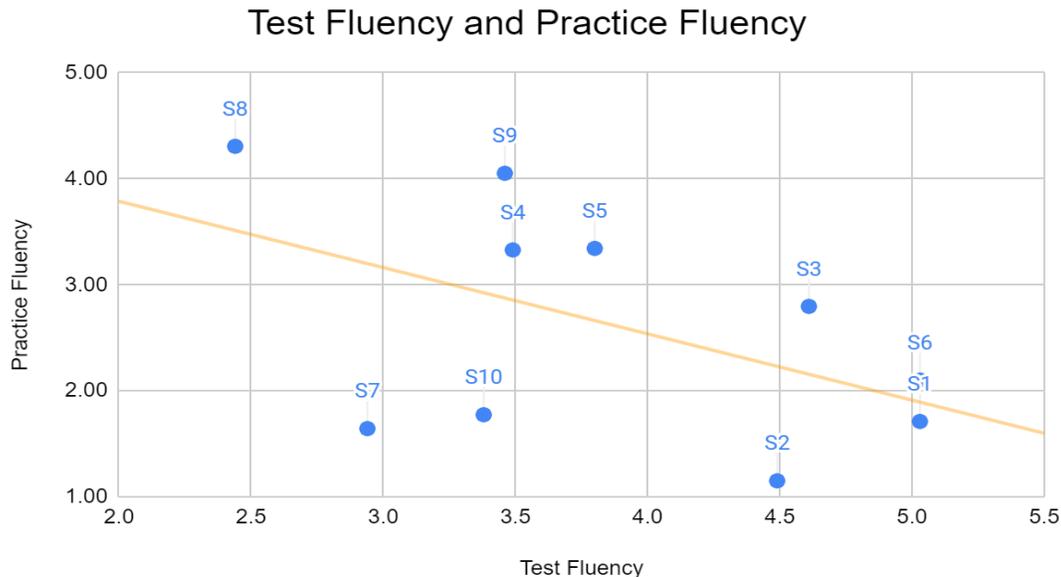


Figure 3: Test Fluency and Practice Fluency

### (4) Practice Frequency, Practice Accuracy, and Test Accuracy

Accuracy is one of the most important factors as well as fluency to have a smooth communication. This section compares participants' accuracy in the Kaizen practice sessions with the accuracy in the speaking test. Figure 4 below shows the comparison between practice accuracy and test accuracy in relation to practice frequency.

In order to measure accuracy in the practice sessions, i.e., practice accuracy, two items were used for analysis. One is the user message count in Kaizen data provided by AB Languages, and the other is the number of sentences learners are required to produce to respond to AI tutors. For example, when the Kaizen user message count recorded five responses in a certain lesson that required only four responses, it was considered that the user made one mistake. The average of the best performance, i.e., the smallest number of responses in each lesson was used as the measure for test accuracy. In other words, if a student practiced seven times in lesson 1, the best performance out of the seven times of practice was used as the test accuracy for lesson 1. The overall accuracy comes from the average test accuracy from lessons 1 through 19. The accuracy in the speaking test, i.e., test accuracy was measured based on the rubric that looked at grammatical and semantic mistakes as well as a communicative skill. Points were deducted per mora or per word for incorrect answers. Figure 4 below shows the comparisons between the practice frequency, practice accuracy, and test accuracy.

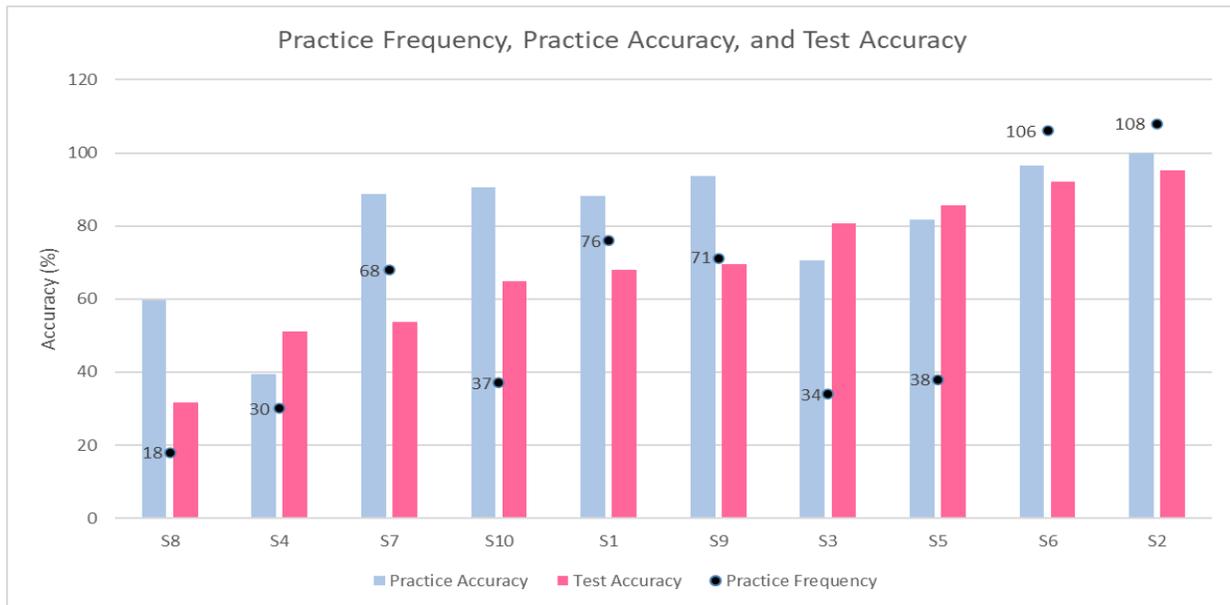


Figure 4: Practice Frequency, Practice Accuracy, and Test Accuracy

Figure 4 shows that students who have a high practice accuracy tend to have a high test accuracy. The figure also shows the more the participants practice, the higher practice accuracy they achieve. The figure also illustrates the similar pattern for the practice frequency and test accuracy whereas some participants such as S3, S5, and S10 show irregularity. For example, S10 has a high accuracy in the practice sessions, but does not have as much high accuracy as the practice accuracy in the test sessions. S10 only completed the first half of the lessons from lessons 1 through 12 one to three times, and thus S10's high accuracy could be attributed to the fact that the participant only had relatively easy practice compared to the other participants who completed more advanced lessons.

The data indicates that S7, S10, S1, and S9 have high practice accuracy without as much practice as S6 and S2. However, the figure shows the mid-range test accuracy for S7, S10, S1, and S9. These results suggest that the participants may be able to improve their practice accuracy only with fewer repetitions while it requires more intensive practice to produce high test accuracy.

In sum, the data suggests that it is necessary for learners to have a number of practice frequencies overall as well as more than two-times of repeated practice of the same lessons in order to apply what they have learned in Kaizen into a new context that requires a spontaneous speech.

### (5) Timing of Fluency Increase

Figures 1 and 2 show that the more frequently learners practice, the more fluent they become. Additionally, the more frequently students repeat the same practice, the more fluent they become. It is not yet known at what point in practice enhances fluency. This section looks at how much repetitive practice is needed to improve L2 oral fluency. To examine the relationship between the number of lessons completed and the time spent in each lesson, the data only includes participants who have completed three or more lessons in a row at least four or more times in each lesson. Three participants were selected for analysis: S1's data covers the performance from lessons 1 through 10, and S2 and S6's data covers the performance from lessons 1 through 19. Figure 5 shows the data for S1's practice frequency and response fluency. Response fluency is the duration of time in seconds it takes for participants to complete a dialogue in each lesson.

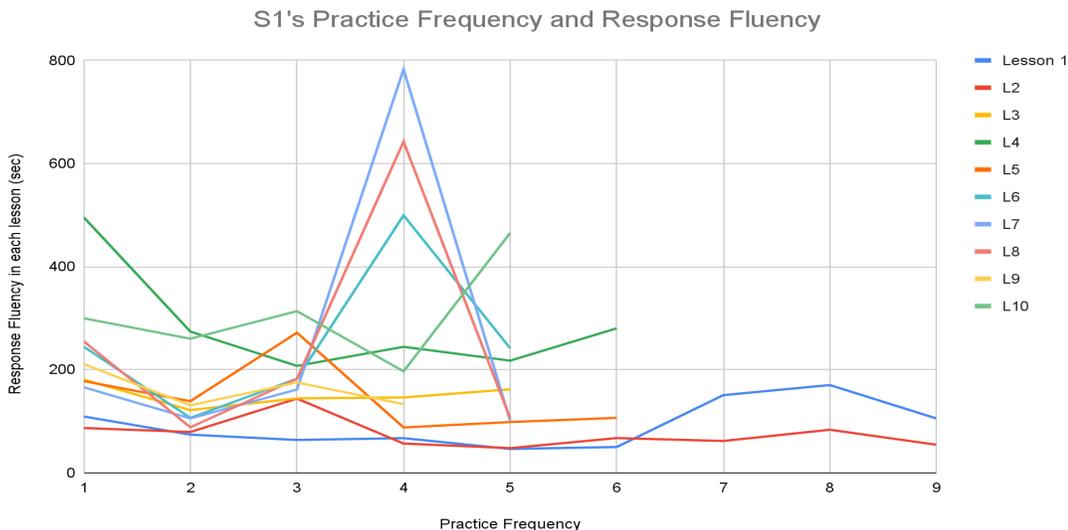


Figure 5: S1’s Practice Frequency and Response Fluency

Figure 5 in x-axis shows practice frequency, while the y-axis shows how long in seconds i.e., response fluency, S1 spent in each lesson. The response fluency in different lessons are displayed in different colors. The response fluency decreases remarkably from the first lesson to the second lesson, and then continues to decrease until about the fourth cycle of practice. The change is not notable after the fourth cycle. Lessons 6, 7, and 8 appear to have a longer duration of time compared to the other lessons, but the fifth practice shows the similar duration of time as the third round, which could be simply caused by participant’s distraction. Figure 6 illustrates S6’s data for lessons 12 through 19. S6 practiced lessons 1 through 19 five times each, except for lesson 18 that had four times of practice. However, due to the overlap between S6’s study period and the time of Kaizen’s system update, the data for the first 2 weeks of S6’s practice was lost. Therefore the performance from lessons 12 through 19 was selected for data analysis. S6 had a large improvement in response fluency from the first to the second round of practice. Such improvement continues until the third and fourth round of practice. The data for S6 shows a similar pattern to S1’s performance.

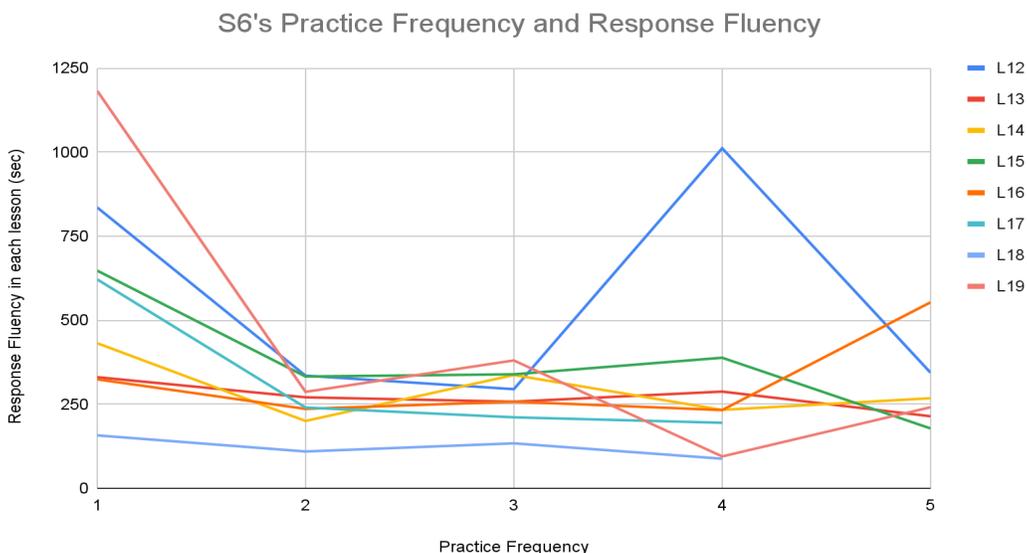


Figure 6: S6’s Practice Frequency and Response Fluency

Only one out of 10 participants completed all the lessons at least five times each, the performance of which is shown in Figure 7. S2 practiced six times in lessons 1 through 7 as well as 12 through 18. S2 also

completed five times of practice in lessons 8 through 11 and 19. The data shows that the response fluency becomes shorter from the first to the second round of practice, and it continues to improve until the third and fourth round of practice. After the fourth time of practice, no notable change is observed in the response fluency, which again illustrates the similar pattern to the results from S1 and S6.

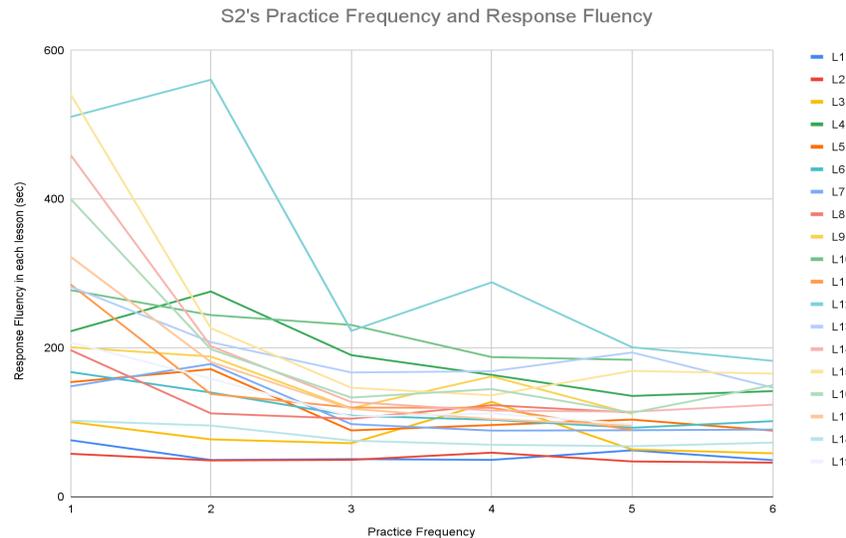


Figure 7: S2's Practice Frequency and Response Fluency

The results from Figures 5, 6, and 7 provide an answer to the research question: *When do learners start improving their fluency? How many repetitions enhance fluency?* The response fluency improved notably from the first to the second round of practice. However, this phenomenon could simply be attributable to the fact that participants had to pay attention to read and understand the instructions as to how to use the app in the first round of practice. In contrast, the second round may have allowed them to focus more on the speaking practice.

The results after the second round of practice were consistent among the three participants, S1, S2, and S6, that the response fluency continued to improve roughly until the third and fourth round of practice. There was not much difference thereafter, which shows more stable change compared to the change from the first to the second round of practice. In sum, at least three or four times of repetitive practice is necessary to improve learners' L2 oral fluency when using a language learning app like Kaizen. The results also confirm that the Kaizen app was one of the useful learning tools to instantly improve learners' oral fluency in a set dialogue. As a future direction, it is important to further examine how many and which lessons learners should cover per day and how frequently they should use the app per week to find more effective learning cycles.

## Conclusion

The study showed that not only did learners' oral fluency improve with the use of the Kaizen app, but their oral accuracy also increased. In addition, learners who practiced repeatedly improved their oral fluency, which required at least three or four times of repetitive practice of the same lesson. Five to six times of practice further produced higher accuracy.

## References

- Ataş, A. H., & Çelik, B. (2019). Smartphone use of university students: Patterns, purposes, and situations. *Malaysian Online Journal of Educational Technology*, 7(2), 54–70.  
<https://doi.org/10.17220/mojet.2019.02.004>

- Blanco, C. (2020, December 15). *The 2020 Duolingo Language Report: Global Overview*. Duolingo Blog. Retrieved April 13, 2022, from <https://blog.duolingo.com/global-language-report-2020/>
- Busuu Revenue and Usage Statistics (2022)*. Business of Apps. (2022, January 11). Retrieved April 13, 2022, from <https://www.businessofapps.com/data/busuu-statistics/>
- Carlson, R. A., Sullivan, M. A., & Schneider, W. (1989). Practice and working memory effects in building procedural skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(3), 517–526. <https://doi.org/10.1037/0278-7393.15.3.517>
- de Bot, K. (1996). The psycholinguistics of the output hypothesis. *Language Learning*, 46(3), 529–555. <https://doi.org/10.1111/j.1467-1770.1996.tb01246.x>
- Duolingo*. BM Education. (n.d.). Retrieved April 13, 2022, from <http://www.bmeducation.in/DUOLINGO/>
- Engber, C. (1989). Applications of cross-language analysis. Karl Sajavaara (ed.). Jyväskylä, Finland: University of Jyväskylä, 1987. pp. 72. *Studies in Second Language Acquisition*, 11(1), 123–123. <https://doi.org/10.1017/s0272263100008068>
- Gatbonton, E., & Segalowitz, N. (1988). Creative automatization: Principles for promoting fluency within a communicative framework. *TESOL Quarterly*, 22(3), 473. <https://doi.org/10.2307/3587290>
- Global EdTech 50: GSV Ventures*. Global EdTech 50 | GSV Ventures. (n.d.). Retrieved April 13, 2022, from <https://gsv.ventures/global-edtech-50/>
- Hamidi, H., & Chavoshi, A. (2018). Analysis of the essential factors for the adoption of mobile learning in Higher Education: A case study of students of the University of Technology. *Telematics and Informatics*, 35(4), 1053–1070. <https://doi.org/10.1016/j.tele.2017.09.016>
- Kormos, J. (2006). *Speech production and Second language acquisition*. Lawrence Erlbaum Associates.
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach\*. *Language Learning*, 40(3), 387–417. <https://doi.org/10.1111/j.1467-1770.1990.tb00669.x>
- Logan, G. D. (1985). Skill and automaticity: Relations, implications, and future directions. *Canadian Journal of Psychology/Revue Canadienne De Psychologie*, 39(2), 367–386. <https://doi.org/10.1037/h0080066>
- Looney, D., & Lusin, N. (2019). Enrollments in languages other than English in United States institutions of Higher Education, Summer 2016 and Fall 2016. *Modern Language Association of America*.
- M., L. W. J. (1989). *Speaking: From intention to articulation*. Massachusetts Institute of Technology.
- McKay, S. (1982). Literature in the ESL classroom. *TESOL Quarterly*, 16(4), 529. <https://doi.org/10.2307/3586470>
- Penglee, N., Christiana, R. W., Battista, R. A., & Rosenberg, E. (2019). Smartphone use and physical activity among college students in health science-related majors in the United States and Thailand. *International Journal of Environmental Research and Public Health*, 16(8), 1315. <https://doi.org/10.3390/ijerph16081315>
- Pew Research Center. (2021, November 23). *Mobile fact sheet*. Pew Research Center: Internet, Science & Tech. Retrieved April 13, 2022, from <https://www.pewresearch.org/internet/fact-sheet/mobile/>

- Roberts, J., Yaya, L., & Manolis, C. (2014). The invisible addiction: Cell-phone activities and addiction among male and Female College students. *Journal of Behavioral Addictions*, 3(4), 254–265. <https://doi.org/10.1556/jba.3.2014.015>
- Rosell-Aguilar, F. (2018). Autonomous language learning through a mobile application: A user evaluation of thebusuuapp. *Computer Assisted Language Learning*, 31(8), 854–881. <https://doi.org/10.1080/09588221.2018.1456465>
- Samuels, S. J. (n.d.). Toward a theory of automatic information processing in reading, revisited. *Theoretical Models and Processes of Reading*, 1127–1148. <https://doi.org/10.1598/0872075028.40>
- Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, 14(04), 357. <https://doi.org/10.1017/s0272263100011189>
- Schneider, W., & Fisk, A. D. (1982). Concurrent automatic and controlled visual search: Can processing occur without resource cost? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8(4), 261–278. <https://doi.org/10.1037/0278-7393.8.4.261>
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84(1), 1–66. <https://doi.org/10.1037/0033-295x.84.1.1>
- Segalowitz, N. (2010). *Cognitive bases of Second language fluency*. Routledge.
- Solomons, L. M. (1899). Communications from the psychological laboratory of Harvard University: Automatic reactions. *Psychological Review*, 6(4), 376–394. <https://doi.org/10.1037/h0074593>
- Survey report on Japanese-Language Education Abroad 2018. The Japan Foundation - Survey Report on Japanese-Language Education Abroad 2018. (n.d.). Retrieved April 13, 2022, from <https://www.jpf.go.jp/e/project/japanese/survey/result/survey18.html>
- Technavio. (2020). *Online Language Learning Market by Product, language, and Geography - Forecast and Analysis 20202024*. <https://www.technavio.com/report/online-language-learning-market-industry-analysis>
- Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, 17(1), 84–119. <https://doi.org/10.1093/applin/17.1.84>
- Wang, M., Shen, R., Novak, D., & Pan, X. (2009). The impact of mobile learning on students' learning behaviours and performance: Report from a large blended classroom. *British Journal of Educational Technology*, 40(4), 673–695. <https://doi.org/10.1111/j.1467-8535.2008.00846.x>
- Wang, Z., & Han, F. (2021). Developing English language learners' oral production with a digital game-based mobile application. *PLOS ONE*, 16(1). <https://doi.org/10.1371/journal.pone.0232671>
- Yoshimura, Y., & Shiomi, Y. (2016) Skritter wo tsukatta onrain zyo deno kanji gansyu no hokoku" [A report of online Kainji learning using Skritter], (Presentation, Annual Conference of Japanese Language Teachers' Association of New England 2016).
- Yoshimura, Y., & Brian, M. (2007) The effect of oral repetition on L2 speech fluency: an experimental tool and language tutor. *SLaTE-2007*, 25-28

Yuen Fook, C., Narasuman, S., Abdul Aziz, N., Syed Mustafa, S. M., & Tau Han, C. (2021). Smartphone use among university students. *Asian Journal of University Education*, 17(1), 282.  
<https://doi.org/10.24191/ajue.v17i1.12622>