Enhancing L2 vocabulary using Anki

Spaced repetition study

Emin Gaaya

Supervisor: Shinichiro Ishihara
Centre for Language and Literature, Lund University
MA in Language and Linguistics, Japanese
SPVR01 Language and Linguistics: Degree Project – Master's (Two Years) Thesis, 30 credits
August 2025

Abstract

This study aims to investigate the effects of the spaced repetition software (SRS) Anki on second language (L2) vocabulary acquisition among learners of Japanese. Over a six-week study, respondents underwent a four-week treatment, learning 150 Japanese–English word pairs using either Anki or their own preferred vocabulary learning strategies. Vocabulary knowledge was assessed through receptive and productive recall in a pre-test, post-test, and delayed post-test format. In addition, respondents' perceptions and experiences were gathered through an interview questionnaire. The sample consisted of learners of Japanese, who were assigned either to an Anki condition ($n = 17$) or a control condition ($n = 17$).

The findings show that the Anki condition significantly outperformed the control condition on the post-test, achieving gains of 45% compared to 33% ($p = .02$, $\eta^2 = .12$). On the delayed post-test, the Anki condition retained 34% of the learned vocabulary, whereas the control condition retained 17% ($p = .03$, $\eta^2 = .10$). In addition, the interviews revealed that respondents had more positive perceptions and experiences with *Anki*, and all expressed a desire to continue using Anki as a supplement to their current vocabulary learning. These findings demonstrate the potential benefits of Anki and spaced repetition software in facilitating both vocabulary gains and long-term retention, as well as their potential for implementation in the classroom.

Keywords: Anki, Spaced repetition, CALL, Expanding spacing, self-study.

Acknowledgements

The author wishes to thank Shinichiro Ishihara for the helpful meetings that enabled me to write this whole study. I want to also thank my respondents for giving me six-weeks of their time, and I apologize if I did not include the data even after helping me for so long. I also want to thank my friends, family and my wonderful girlfriend for helping me and keeping me on track.

Table of Contents

List of figures

List of tables

Convention

Japanese vocabulary items are written in hiragana, with the English translation provided in parentheses.

Abbreviations

SRS	Spaced Repetition Scheduler / software

1. Introduction

This study investigates the effectiveness of Anki, a spaced repetition software (SRS), for enhancing Japanese vocabulary retention among second language learners. Achieving proficiency in a second language (L2) requires learners to remember vocabulary long-term. They must master thousands of words, from individual to multi-word expressions, to support compelling reading, listening, writing, and speaking (Nation, 2006). To achieve this, deliberate vocabulary practice, such as using flashcards (word cards), is one of the most effective strategies for establishing meaning from connections (Nakata, 2019; Webb et al., 2020). Flashcards are usually physical paper with a question written (L1 word) on one side and an answer (L2 word) written on the other (Mondria & Mondria-de Vries, 1994; Nation, 2001, p. 296). This is a form of paired-associate learning that, when used correctly, helps strengthen connections between two items and supports faster vocabulary acquisition (Thorndike, 1908; Fitzpatrick et al., 2008; Steinel et al., 2007; Nakata, 2011). However, even if flashcard practice is effective, retention tends to decay and therefore requires reinforcement (Nakata, 2019, p. 308). To achieve long-term retention, using flashcard studying with spaced repetition is one of the most effective ways to practice new vocabulary. The current study discusses only relative spacing, which can be divided into two spacing methods. First is equal spacing, where reviews occur at consistent intervals; for example, reviewing material every four days and expanding spacing, where the intervals between reviews gradually increase, such as reviewing after three days, then seven days, and later after nine days (Nakata, 2011, p. 201). Both methods are optimal to increase long-term retention. However, equal spacing is perhaps not as realistic as learning a new language, as adding new words constantly and studying new vocabulary every fourth day is a workload no one can handle (Schuetze and Weimer-Stuckmann, 2010, 2011). For those reasons, expanding is perhaps more optimal as it pushes old cards further away and lets new cards be added (Bjork, 1988, p. 399; Mondria & Mondria-de Vries, 1994). However, the researchers do not agree on what spacing method is the best, but they all agree that spacing is better than massing, a method in which a person studies everything at once, with no spacing between study sessions. For example, having one continuous study session of three hours instead having three study sessions of one hour each spaced along the day. Spacing seems to always win against massing in this aspect.

The present study is a six-week study that investigates the use of Anki, a spaced repetition software (SRS), a program that includes flashcards and expanding intervals into its review system. Anki builds upon the Leitner system (Leitner, 1972), a traditional approach that organizes flashcards into boxes. When a card is answered correctly, it advances to a longer review interval, for example from one day to four days. When a card is answered incorrectly, it returns to a shorter interval, for example from four days back to one day. Unlike the fixed structure of physical flashcards, Anki employs an adaptive algorithm that ensures difficult items reappear before the interval becomes too long, while familiar items are not reviewed excessively. In contrast to physical flashcards, which are bulky, require storage space, and lack portability, digital spaced repetition software offers a more practical alternative with greater flexibility and require only an initial internet connection before can be used fully offline. Examining the use of SRS in authentic learning contexts, rather than exclusively in controlled laboratory environments, is crucial for producing realistic and pedagogically meaningful findings.

The gap in the research comes from the fact that much of the existing research has been conducted in controlled laboratory settings, which may not accurately represent authentic language learning conditions. This study addresses this limitation by using a realistic approach in a six-week study involving 34 learners of Japanese. Respondents used the spaced repetition software (SRS) Anki or a self-selection vocabulary learning strategy that was not Anki in a control condition to learn 150 English-Japanese word pairs. The effectiveness of their learning was measured through a pre-test and post-test format, with an additional delayed post-test to assess retention. Additionally, two interviews via Google Forms used to collect data on motivation and the opinion within each condition.

The purpose of this study is therefore to examine the effectiveness of Anki in comparison to self-directed study methods. Specifically, it seeks to answer the following research questions:

1. Is there a significant difference in vocabulary gains between digital flashcards using the spaced repetition software Anki and the control condition?
2. Is there a significant difference in vocabulary retention between digital flashcards using the spaced repetition software Anki and the control condition?
3. What are the respondents' perceptions and experiences regarding the effectiveness of vocabulary learning strategies in the Anki condition compared to the control condition?

The structure of this study begins with an introduction, followed by a literature review that summarizes previous research findings and emphasizes the theoretical framework of spaced repetition. This section also includes an analysis of prior studies on Anki. The third section presents the research question, while the fourth describes the methods and materials. The fifth section reports the results, which are then discussed in the sixth section. The study concludes with references and an appendix.

2. Literature Review

Vocabulary knowledge is widely regarded as the most essential component of language acquisition. While grammar is important, communication cannot occur without vocabulary (Schmitt, 2008). Among the various tools used to study vocabulary, flashcards are especially common. In Japan, for example, 60% of students report using flashcards; however, many employ ineffective strategies (Nakata, 2011; Zung et al., 2022). Flashcards are effective because they promote retrieval practice, which has been shown to substantially enhance retention (McDaniel & Fisher, 1991). When combined with spaced repetition, these benefits can become even more pronounced (Cepeda et al., 2006).

The aim of this literature review is to examine existing research on spaced repetition and retrieval practice, particularly as implemented in computer-based flashcard programs. Special attention will be given to the underlying theories supporting these methods, with a focus on demonstrating the benefits of such tools and analyzing the science behind *Anki's* spaced repetition system. I will begin by going over early research on spaced repetition, then explain studies on spacing, computer assisted language learning (CALL), flashcards.

2.1 Early research on spaced repetition

When we aim to learn something new, we often encounter moments when inevitable information slips from our memory. Forgetting is a natural aspect of being human, and despite our best efforts, it is something we cannot entirely avoid (Ellis, 1995; Hulstijn, 2001; Nation, 2001). However, many have experienced remembering something we had forgotten, experiencing a stronger connection the second time but nonetheless forgetting it again the third time. The concept of forgetting can be abstract and complex to grasp, but research by Ebbinghaus in 1885 began to shed light on this topic. Research by Ebbinghaus in 1885, marks the beginning of systematic, empirical research on memory. His pioneering work laid the groundwork for understanding how we acquire, retain, and eventually forget information. One of his most significant contributions was the development of the "forgetting curve," a graphical representation that demonstrates the rapid decline of memory retention following the initial learning of new information (Ebbinghaus, 1885; Mondria & Mondria-de Vries, 1994).

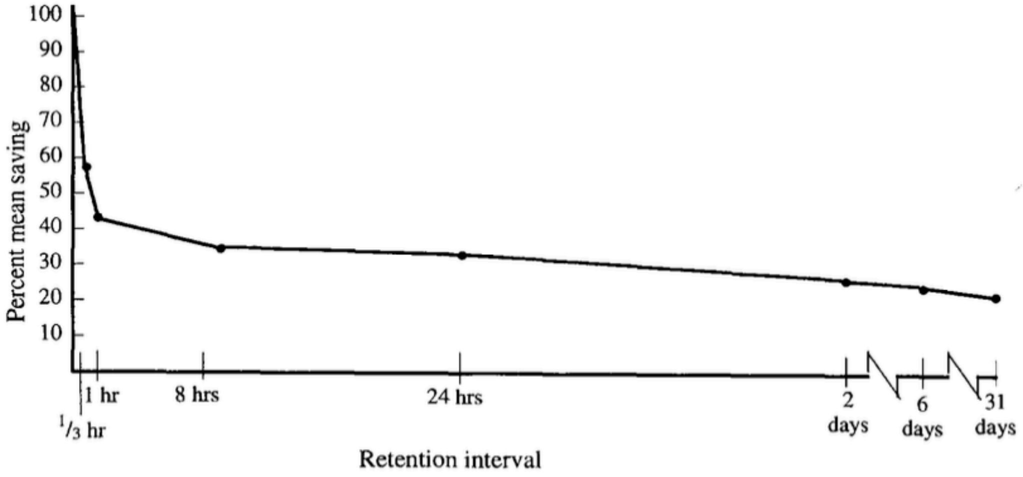


Figure 1 Forgetting curve (Ebbinghaus,1885; Mondria & Mondria-de) Vries, 1994).

According to Ebbinghaus, forgetting is not a flaw in our learning system but an inherent part of the memory process. He noted that when we are on the verge of forgetting newly acquired information, reviewing it at that critical juncture can significantly aid retention (see figure 1). For instance, when you learn new information, the curve shows that the initial forgetting is around 60-70% during the first 24 hours. However, if you review this information the following day, the memory trace is strengthened, and it remains accessible for a longer period before being forgotten again. As you continue to review the information, the rate at which it is forgotten decreases. This is because each subsequent review reinforces the memory further, causing the forgetting curve to flatten. A flatter curve indicates that the information takes longer to be forgotten, making it more likely to be retained in long-term memory. In other words, with each review session, you are effectively extending the period during which the memory remains accessible. Over time, the intervals between reviews can be lengthened, reflecting the fact that the reinforced memory is now more robust and resistant to decay. Ebbinghaus lays the empirical foundation for modern spaced repetition system and research on forgetting. The method most closely associated with Ebbinghaus use of nonwords and seeing how fast these words were forgotten (see Murre & Dros, 2015, for an updated replication and analysis of the forgetting curve).

Further expanding on this research, Pimsleur (1967), best known for his language learning program, developed guidelines on how quickly information should be reviewed to reach 100% recollection (see Figure 2), following the same principles of the forgetting curve. Pimsleur proposed that the optimal moment to review material is when approximately 40% has been forgotten. With each review, the memory is reinforced, allowing for longer intervals before the next review. Over time, these intervals expand to the point where the information can be retained without additional review.

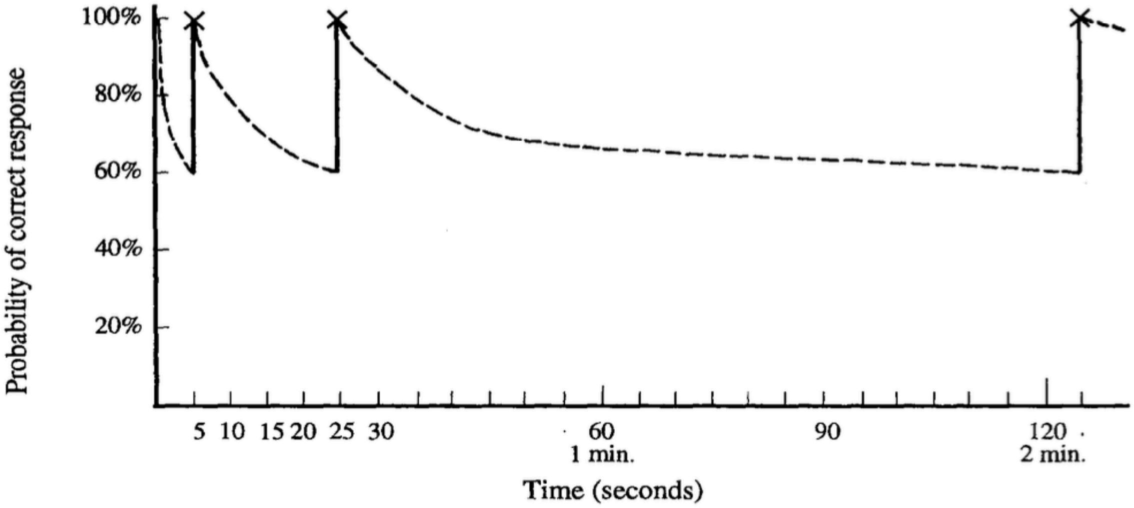


Figure 2 Pimsleur Forgetting curve (Mondria & Mondria-de Vries, 1994).

More research on long intervals has been studied by both Lado (1964) and Landauer and Bjork (1978), which started the push for Pimsleur expanding spacing. The reviews are seen to improve long-term retention, and according to Lado, even after each review, two- or three-times amount of the original interval is needed until the next review (Mondria & Mondria-de Vries, 1994, p. 50). However, one should only expand spacing between reviews if someone recalls the correct information, and there should be a system the learner can follow if they do not recall incorrectly.

One effective learning system is the Leitner system, developed by Sebastian Leitner in 1972. This flashcard organization method lets learners to schedule their review sessions using expanding spacing more efficiently. The Leitner system consists of a box with five compartments (see figure 3).

To begin using this method, learners must first create their flashcards. As Nakata (2011) describes, Flashcards (word cards) are usually a piece of paper with typically an L2 word on one side and its meaning on the other, usually provided as a translation in the learner's first language (L1). However, these flashcards can be definitions and explanations. There could also be an equation of one side and the solution on the other (Polly et al., 2025).

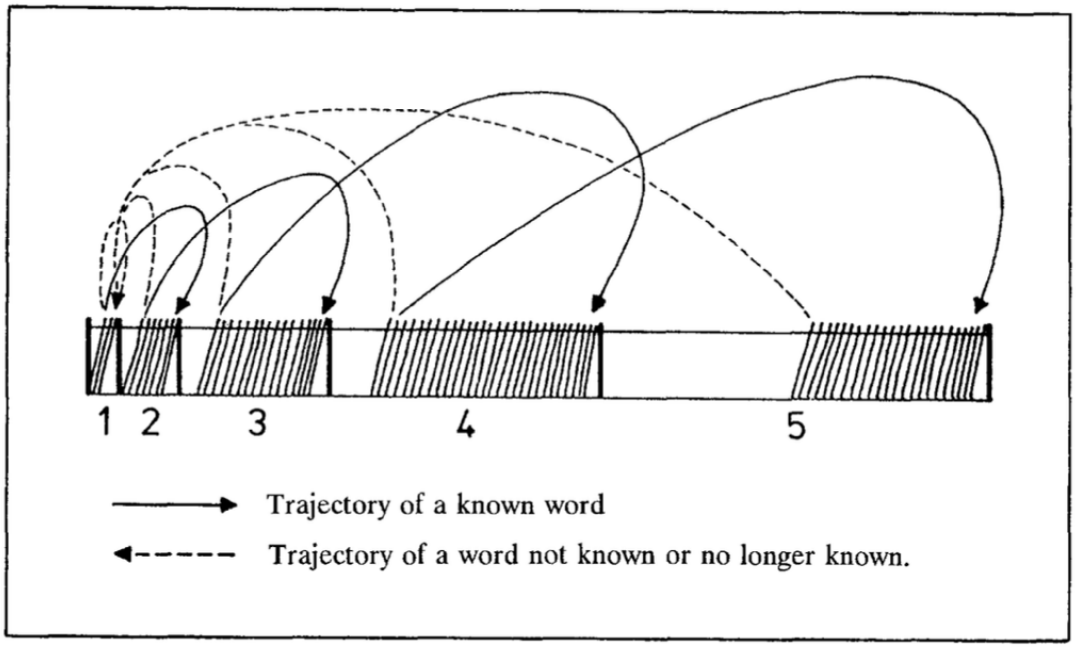

1
2
3
4
5
Trajectory of a known word
Trajectory of a word not known or no longer known.

Figure 3 Leitner's hand computer, (1972; Mondria & Mondria-de Vries, 1994, p. 52).

In the first session, the learner starts with approximately 30 to 40 cards (Mondria & Mondria-de Vries, 1994, p. 52). They examine both sides of a flashcard that displays a word in their native language on one side and its translation on the other. The learner goes through all the cards until they repeat the first card. If they recall the card correctly, they move it to the second compartment. The learner continues studying the first compartment until they can recall all the cards accurately.

When learners correctly recall a word from the second compartment, they graduate it to the third one. However, if they recall it incorrectly, the card returns to the first compartment. The learner reviews and graduates known cards as compartments fill up while returning incorrectly recalled ones to the first compartment. When the final compartment is complete, the learner can discard the cards as understood or store them, with occasional reviews to ensure they retain the information.

Combining these two principles forms the basis of many spaced repetition software programs, such as *Anki*, *SuperMemo*, *Memrise*, and *Quizlet* (see Nakata, 2015, for a review). These digital tools also address some of the limitations of the physical method, such as the serial position effect and the physical method can be bulky and take up much space, especially depending on how many cards are added. A few studies on Leitner's hand computer (Leitner's learning box) have been conducted using the method more recently. First is Farhadi (2012), who tested the Leitner's learning box compared to a control condition on a vocabulary study with a pre-test and a post-test showing the of Leitner's hand computer outperformed to the control condition. Another study, Whitmer et al (2022), implemented the method in a digital lab setting and saw the Leitner's hand computer to be more efficient in study time but did not show better retention.

2.2 Theoretical background

Previous studies have investigated different forms of spacing and its effect on retention. The current study investigates spaced repetition software Anki, which uses expanding spacing with the incorporation of Leitner's learning box to try to help learners remember longer with digital flashcards. The study takes early research on spaced repetition and creates a program around this method. However, spacing has shown great significance on retention compared with different methods.

Spacing can be divided into two forms: absolute spacing and relative spacing. Absolute spacing refers to the total amount of time between the first and last study session. For example, if someone studied a word three times with a 2-day interval between each session, the absolute spacing would be 6 days (Karpicke & Bauernschmidt, 2011; Sonbul et al., 2024).

In contrast, relative spacing concerns the pattern or distribution of intervals between study sessions. Relative spacing can be further categorized into two types: equal spacing and expanding spacing (Pyc & Rawson, 2007; Karpicke & Schmidt, 2011; Kang et al., 2014; Nakata, 2011). Equal spacing involves reviewing material at consistent intervals, such as every 4 days. In expanding spacing, the intervals between reviews increase over time, for example, reviewing a word after 3 days, then after 5 days, and then after 9 days (Karpicke & Bauernschmidt, 2011).

Another important factor influencing the effectiveness of spacing is the lag effect, which refers to the impact of the length of intervals between study sessions. For instance, if someone studies over a total period of 12 days with either equal intervals of 3 days (e.g., days 0, 3, 6, 9, 12) or expanding intervals like days 1, 4, and 7, the absolute spacing in both cases would still be

12 days, but the relative spacing differs. The lag effect shows that studying with longer spacing between sessions generally leads to better long-term retention, whereas studying with shorter intervals results in better short-term retention (Cepeda et al., 2006; Nakata, 2011; Karpicke & Bauernschmidt, 2011; Nakata et al., 2023). The phenomenon is called the spacing effect, which empirical research has proven leads to better long-term retention when one spaces out their study sessions over time, rather than studying every day with little to no gap between sessions, which is called massing (Feng et al., 2019; Yamagata, et al., 2023). Additionally, when one uses massing right before an exam that is a common term called cramming which is what massing is, but it does not necessarily mean a test is common. Massing is a common strategy that may help with short-term gains, which explains its popularity for exam preparation, but it does not support long-term retention. (Nakata, 2011). However, right after the tests all the information disappears fast unless it is reviewed again which will turn it into spacing, but people who use massing do not normally return to the literature and instead have one long session of studying. However, cramming is probably the most common way of cramming as pure massing is not that common (Cepeda et al., 2008; Kornell, 2009).

Previous studies has yet to prove the best form of spacing between equal and expanding spacing. Studies have shown results that equal spacing to be better at post-test scores (e.g. Pyc & Rawson, 2007; Storm et al., 2010). However, studies have also shown an advantage for expanding spacing (Vlach et al., 2014; Nakata, 2015; Kanayama, 2020). Although, studies on relative spacing and massing, shows spacing has an advantage over massing (e.g. Zulkiply, 2013; Sonbul, 2024). Furthermore, studies on language learning have shown statistical advantage of spacing (e.g. Verkoeijen et al., 2008; Rogers, 2015; Kim & Webb, 2022; Yan & Zhou, 2023)

Nakata (2015) noted that second language (L2) vocabulary acquisition consistently shows that spaced repetition is more effective than massing (Cepeda et al., 2006; Nation, 2001). Within spaced repetition, there are two major scheduling approaches as mentioned before: equal spacing and expanding spacing. Although expanding spacing has received considerable empirical attention, results have been mixed, and no study has conclusively shown that expanding spacing is more beneficial than equal spacing in terms of test scores (Karpicke & Bauernschmidt, 2011; Pyc & Rawson, 2007). Some studies have found that direct or delayed feedback combined with equal spacing yields significant results (Logan & Balota, 2008; Storm et al., 2010). However, Nakata (2015) reported that expanding spacing produced slightly better results compared to equal spacing in a realistic second language paired-associate learning setting that included productive recall and immediate feedback. Further research is needed to explore these findings, but the overall literature indicates that spaced repetition is superior to massing for L2 vocabulary acquisition.

Another study by Nakata & Suzuki (2019), which investigated the effects of massing compared to spaced repetition. The study focused on vocabulary retention by examining both semantically related and unrelated words. The sample consisted of 133 Japanese university students who had received at least six years of English language instruction. Respondents were divided into two conditions. One condition used a massing approach and studied semantically related or unrelated words consecutively, while the other condition employed spaced repetition, with repetitions of the same words distributed over time. The study materials comprised 48 low-frequency English words paired with their Japanese translations. The words were carefully matched for factors such as frequency, length, and familiarity. Half of the words were semantically related, for example, conditions of animals or plants, while the other half were unrelated. Recall of the Japanese translations were assessed in both immediate and delayed post-tests to investigate long-term retention. The results demonstrated significant benefits from spaced repetition. In addition, semantically related words produced more recall errors than unrelated words, suggesting that semantic clustering may hinder retention.

Additionally, research has explored various techniques within spacing. Studies have examined the benefits of repeated reading (Serrano & Huang, 2018), differences in testing formats (Nakata et al., 2021), the effects of spacing on vocabulary acquisition (Yan & Zhou, 2023), and comparisons between study materials such as flashcards and word lists (Nakata, 2008). Further investigations in classroom settings have also demonstrated the superiority of spaced repetition for learning (Rogers & Cheung, 2020). Nakata (2008), in particular showed that computer-based flashcards significantly outperform both physical flashcards and using a word list. Spacing or spaced repetition as it is often called has been shown to be beneficial or learning information long-term. Taking this information, we might find ways to be able to study more with programs and software that aid the spacing process.

Vocabulary knowledge

The current study employs both productive and receptive recall tasks in its pre- and post-tests. According to Nation (2001), productive knowledge refers to the ability to actively use vocabulary in tasks such as speaking and writing, while receptive knowledge involves recognizing and understanding vocabulary through listening and reading. Prior research has shown that receptive knowledge is generally easier to acquire than productive knowledge, as it relies on recognition and is closely tied to meanings already established in the learner's first language (L1). In contrast, productive knowledge requires greater retrieval effort, as learners must retrieve the second language (L2) word which require more effort (Nation, 2001; Yanagisawa, 2016).

Learning vocabulary with sound

The current study employs the Anki flashcard program to enhance vocabulary learning by incorporating audio. Although research on the use of sound in vocabulary acquisition remains limited (e.g., Chun & Plass, 1996; Nation, 2001; Plass & Jones, 2005; Teng, 2023), and no prior work has examined user-created flashcards with multimedia elements. However, Teng's study provides a helpful insight on the benefits of multimedia input in language learning. In his experiment, respondents were assigned to one of four conditions: "definitions alone, definitions with detailed word information, descriptions and word information accompanied by audio or definitions, and word information accompanied by a video", and their retention was measured with a delayed post-test administered two weeks after the initial treatment. The findings showed that all conditions outperformed the ‘’definition’’ only condition. However, the study did not support sound to be the most significant in the results it was the second-best factor on retention on all the conditions in the study (p. 747). Teng (2023) mentions that dual coding theory may explain why retention was higher in the sound and video condition. This theory, introduced by Allan Paivio in 1986, proposes that verbal and non-verbal processing occur simultaneously, which enhances working memory and ultimately leads to better long-term retention (Teng, p. 739).

2.3 Space repetition software (SRS) and flashcard programs

Numerous studies have compared the effectiveness of computer-assisted language learning programs with the incorporation of spaced repetition schedules, showing their benefits over traditional methods (Cakmak et al., 2021; Bower and Rutson-Griffiths, 2016; Yüksel et al., 2022). Traditional benefits can be defined as any method that do not include digital software. Such as, Pen and paper, physical flashcards, but also reading and taking notes.

Some studies also compare ways digital programs are better than others. With the popularity of various apps such as Duolingo and Quizlet, decided to compare alternative methods to see if one is better. For example, Larchen et al. (2020), investigated Quizlet and a virtual reality program to study 10 idioms. The study tested two spacing conditions 15 min and 1 week and then tested their benefits on a one-week retention post-test. The study concluded that the virtual reality program produced better retention across both spacing conditions. Another study by Jia et al. (2023) found that, on a three-week delayed post-test, respondents showed greater long-term retention when using the program compared to Quizlet and a paper-based flashcard game. A further study involving Quizlet, conducted by Bueno-Alastuey and Nemeth (2020), compared podcasts with Quizlet for receptive and productive vocabulary knowledge. However, this study reported no significant difference between the two conditions.

2.4 Evaluating Spaced Repetition Software Anki

2.4.1 Anki in the medical field

Kaitsu and Nakata (2025) proposed comprehensive criteria for evaluating computer-assisted language learning (CALL) software (2011) and mobile-assisted language learning (MALL) applications (2025). Both papers assess widely used programs such as Quizlet, SuperMemo, and iKnow!; however, Anki, the program used in the present study, was not evaluated. According to Kaitsu and Nakata (2025, p. 3), this omission was due to prior evaluations conducted by Koleini et al. (2024) and by Dunlosky and O'Brien (2020), though those evaluations addressed only the desktop version of Anki. To address this gap, we will apply Kaitsu and Nakata (2025) mobile criteria to evaluate both the mobile and desktop versions of Anki. We chose this approach because the desktop criteria Nakata established in 2011 are similar to those from 2025 and improved. Furthermore, unlike Dunlosky and O'Brien (2020), who focused on evidence-based learning strategies rather than language learning per se, Kaitsu and Nakata's frameworks are targets programs used to learn languages.

Dunlosky and O'Brien (2020) examined whether Anki and other spaced-repetition programs support effective learning strategies such as spaced retrieval and successive relearning (SR). A program that implements these strategies well can improve students' performance regardless of background or field of study (Dunlosky & O'Brien, 2020, p. 227). Their study employed a thirteen-criterion evaluation framework to assess the effectiveness of spaced-repetition software. Anki scored near-perfectly across all thirteen criteria, including the ability to add images, customize scheduling options, and accept typed answers. However, Dunlosky and

O’Brien (2020) presented in their results that Anki lacks built-in support for features such as multiple-choice testing and automated external reminders (e.g., email or SMS notifications), which must be enabled via add-ons or user configuration via code. For users without technical expertise, Anki’s large community offers a wide range of pre-made decks (Set of flashcards) and third-party add-ons available on AnkiWeb. In sum, Dunlosky and O’Brien conclude that Anki effectively supports successive relearning and enhances learning outcomes, though the specific version evaluated is not clearly stated (presumably the desktop client, possibly alongside the web version). Future research should specify the exact Anki version under investigation, as functionality can differ substantially across platforms.

Koleini et al. (2024) explicitly evaluated the mobile (application) version of Anki. Their study assessed the acquisition of 100 technical vocabulary terms by 80 Iranian university students majoring in Psychology over a ten-week treatment period, comparing mobile-assisted digital flashcards (Anki) with traditional paper flashcards. Using the Vocabulary Knowledge Scale (VKS) as a pre-test, immediate post-test, and six-week delayed post-test, they analyzed results via a 2×3 mixed-design ANOVA. Respondents studied for fifteen minutes per day, Monday through Friday. Although Anki’s built-in statistics tracked total study time and number of cards reviewed, these metrics were not reported. The digital-flashcard condition significantly outperformed physical flashcard condition on both immediate and delayed post-tests. However, the study relied solely on self-reported data, without objective usage metrics, satisfaction surveys, or qualitative interviews, commonly included in another medical-education research. For instance, Harris and Chiang (2022) and Jape et al. (2022) report high user satisfaction with Anki among medical students, and Wothe et al. (2023) found improved sleep quality in Anki users.

Goldman et al. (2024) conducted a systematic review of eight studies published between 2015 and 2022, demonstrating a positive correlation between Anki usage and significant improvements in exam scores (e.g., Deng et al., 2015; Wothe et al., 2023; Lu et al., 2021; Gilbert et al., 2023; Levy et al., 2023) as well as a lower failure rate (2.8% vs. 10.94%; Cooper et al., 2023). Strauss, et al., (2019) similarly found that residents using Anki achieved a 92% pass rate, well above the national average of 67%. However, some studies (e.g., Sun et al., 2021; Levy et al., 2023; Cooper et al., 2023) report no significant score improvements compared to alternative methods. These discrepancies may stem from differences in Anki implementation, user engagement, or enjoyment levels.

Although studies focusing on *Anki* in language learning are limited, it is essential to review existing research that has examined its effectiveness, as the present study directly addresses this gap. For example, Indonesian second-grade students (Jaya, 2020) and adult EFL learners (Iravi & Malmir, 2023) demonstrated significant post-test gains following Anki-based practice, while college-level ESL students also improved their vocabulary exam scores (Ozer et al., 2017). Additionally, Iranian learners who received Anki-based instruction retained more new words than students in traditional classes (Khoshsima & Khosravi, 2021), and Indonesian vocational students achieved significant Japanese-vocabulary acquisition using the AnkiDroid app (Nender et al., 2022). Furthermore, Anki studies has shown after 10 weeks treatment of with a long-term retention of Spanish vocabulary compared to the non-Anki method (Mujahidah et al., 2024), and university students attained higher end-of-semester test scores after Anki treatment (Seibert Hanson & Brown, 2019). Based on results from spaced repetition software one can theorize that

Anki in the current study will both show better delayed post-test performance over the methods in the control condition similar to the other studies on spaced repetition software Anki.

2.4.2 Evaluating the Flashcard program Anki

We evaluate Anki using the framework proposed by Kaitsu and Nakata (2025). Our primary focus is the desktop client, which offers the most extensive functionality. Particularly because of the third-party add-ons. Nevertheless, we also examine the mobile version of Anki, because respondents in the current study use both platforms. Even if the present study does not exploit every available feature, it is still important to assess Anki's overall suitability for language learning.

Kaitsu and Nakata (2025) identify twenty-four criteria for assessing mobile flashcard software. Nakata's earlier framework (2011), developed for desktop flashcard programs, addressed many of the same aspects, but the 2025 revision both refines existing criteria and introduces new ones. The updated framework provides a more comprehensive tool consisting of seven criteria for flashcard creation and editing and seventeen for learning benefits (p. 12).

Following Kaitsu and Nakata's scoring system, Anki will be evaluated in the same way. A plus sign (+) indicates that a criterion is met and earns one point; a minus sign (–) indicates the criterion is not met and earns zero points; and "N/A" is used for features requiring programming, also earning zero points. A double plus (++) denotes that a feature goes beyond the baseline expectation but is still counted as one point. For example, criterion 16 (block size) receives a double plus when users, rather than the program, are free to set the block size themselves (p. 13).

According to Kaitsu and Nakata (2025, p. 5), a flashcard program should provide learners with a library of pre-made decks so they can begin studying immediately. However, research also shows that creating one's own flashcards leads to stronger retention (Dodigovic, 2013; Lei & Reynolds, 2022). Therefore, an effective program must combine both functions: offering ready-made decks for accessibility while also allowing learners to generate their own cards to maximize learning outcomes. Further, it should also support fully multilingual input (alphabetic and non-alphabetic) so that users can study any language, which is a clear advantage when writing in the L2 during learning to enhance acquisition (Gyllstad et al., 2023). Additionally, the software should accommodate multi-word items (e.g., idiomatic expressions), which contribute to greater fluency in the target language (Schmitt, 2023).

Kaitsu and Nakata (2025) also recommends features such as the ability to organize cards into sets for easier categorization and to share those decks with other learners and instructors, thereby facilitating feedback and collaboration (Dunlosky & O'Brien, 2022). Finally, incorporating multimedia elements accords with dual-coding theory, which posits that pairing verbal and visual information strengthens retention (Paivio & Desrochers, 1980); empirical evidence likewise supports the use of images in vocabulary learning (Carpenter & Olson, 2012; Ramonda, 2022). According to this creation and editing criteria, Anki scores a perfect seven out of seven. Users benefit from an extensive community-curated collection of decks hosted on AnkiWeb, covering a wide range of languages (e.g., Japanese, French) and subjects (e.g., physics, music). The same advantages apply to the mobile apps.

The major difference between the desktop and mobile versions is that the mobile apps do not support user-created add-ons. Consequently, on mobile. Kaitsu and Nakata's last 17 criteria focused only on learning and Anki mobile fails to meet 6 out of 17 of them. Anki desktop can add this function with the help of add-ons or coding. Kaitsu and Nakata's criteria 10 (receptive recognition), 12 (productive recognition), 14 (varied encounters and use), 18 (fluency development), 19 (automatic speech recognition), and 24 (motivational feedback). As discussed in the literature review, productive and receptive recall are more beneficial than recognition, but an effective flashcard program should support all formats according to Kaitsu and Nakata. Criterion 14, varied encounters and use, refers to exposure to words in multiple contexts (e.g., several example sentences) to deepen lexical knowledge (Kaitsu and Nakata, 2025, p. 7). Fluency development (criterion 18) describes progression from beginner to communicative competence.

Furthermore, the mobile apps do meet criteria 11 (receptive recall) and 13 (productive recall), which lie at the core of Anki's functionality. Anki also satisfy criterion 17 (interference avoidance) by allowing users to study specific tagged cards or decks. Criterion 20 (adaptive sequencing) is met through Anki's spaced-repetition algorithm. Criterion 23 (formative feedback) is not provided automatically but can be added by users or generated through add-ons. Anki's direct feedback lets learners include as much information as they need for later review. Motivational feedback (criterion 24) is available via add-ons such as Anki Leaderboard, which introduces gamification elements that enhance motivation (p. 10).

Kaitsu and Nakata excludes criteria 21 (retirement) and 22 (expanding retrieval) from the overall score due to limited empirical evidence regarding their optimal implementation (p. 14).

Applying his scoring to the mobile version of Anki, alone receives 15.5 points out of 46. This is one of the lowest scores if we compare it to the already evaluated programs evaluated by Kaitsu and Nakata. However, with the combination of the desktop version the Anki version gains the same capabilities as the desktop version. Further if looking at the mobile version alone Anki mobile still ,employs recall retrieval, which the given study uses. When the desktop version with add-ons is considered, Anki's score rises to 45 out of 46. These results reflect Anki's flexibility: it imposes no limitations and enables users to create fully customized study environments. This openness has driven its widespread adoption in medical and language-learning contexts. Although Anki may initially seem challenging, requiring basic coding for full functionality. The large user community, free add-ons, and vast library of pre-made decks make it as user-friendly as out-of-the-box applications such as Quizlet and Memrise.

3. Research Questions and hypothesis

The purpose of this study is to evaluate and compare the effectiveness of the spaced repetition software (SRS) *Anki* with other study methods in terms of vocabulary acquisition and retention when learning 150 Japanese–English word pairs. The study investigates whether *Anki* can be implemented effectively and whether respondents using it are able to learn the vocabulary within a four-week period. The current study takes a more realistic approach by letting a control condition choose when they study and how they study. Anki is a popular in the language learning community and shows its effectives in previous research in the medical field and in some limited studied in language learning.

The current study based in previous research creates two hypothesis that are based on the previous research and the potential advantages and disadvantages spacing has over other methods.

H1: Anki will lead to better long-term retention of 150 Japanese-English word pairs on the two-week delayed post-test compared to a control condition.

H2: The control condition will lead to better post-test scores of 150 Japanese-English word pairs on the four-week study period compared to the Anki condition.

The Anki condition is expected to perform worse than the control condition on the post-test. Previous research has shown that while spacing is effective for long-term retention, massed practice tends to produce better results for short-term retention.

The current study wants to answer three research questions

RQ1. Is there a significant difference in vocabulary gains between digital flashcard using the spaced repetition software Anki compared to the control condition.

RQ2. Is there a significant difference in vocabulary retention between digital flashcard using the spaced repetition software Anki compared to the control condition.

RQ3. What are the respondents' perceptions and experiences regarding the effectiveness of vocabulary learning strategies in the control condition compared to the Anki condition?

4 Methodology

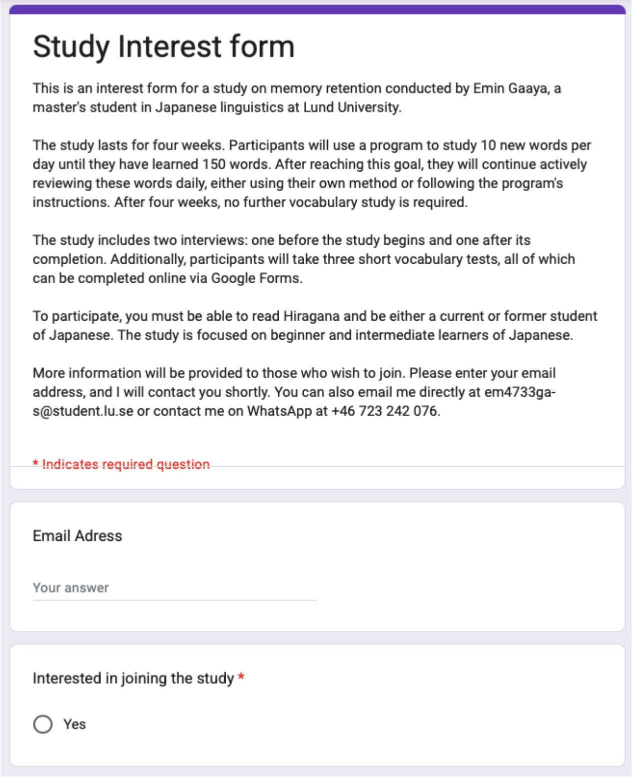
This section begins by reviewing how we recruited and what criteria were used to select the respondents for the study. Next, we describe the materials and instruments, explicitly focusing on vocabulary selection and wordlist creation. We then explain the flashcard program, Anki. Finally, we provide a detailed procedure outline and explain how we gathered and analyzed the data.

4.1 Respondent

The recruitment phase took place during February and March 2025. The researcher contacted universities in Sweden, Norway, Denmark, and Finland that offer Japanese as a second language. To attract respondents, four digital Zoom information sessions were held by the researcher, in addition to two in-person sessions on campus.

Fliers containing a QR code were distributed on campus, on social media, and in online forums for Japanese learners. One such forum was Reddit, which hosts a community dedicated to Anki with approximately 169,000 members, including many learners of Japanese. Interested individuals were invited to register their interest either by email or through a Google Form. The QR code redirected respondents automatically to the researcher's email, while the Google Form led them to the information page displayed in Figure 4. Multiple versions of the flier were developed to attract new respondents. The final version is presented in Figure 5

.



Study Interest form

This is an interest form for a study on memory retention conducted by Emin Gaaya, a master's student in Japanese linguistics at Lund University.

The study lasts for four weeks. Participants will use a program to study 10 new words per day until they have learned 150 words. After reaching this goal, they will continue actively reviewing these words daily, either using their own method or following the program's instructions. After four weeks, no further vocabulary study is required.

The study includes two interviews: one before the study begins and one after its completion. Additionally, participants will take three short vocabulary tests, all of which can be completed online via Google Forms.

To participate, you must be able to read Hiragana and be either a current or former student of Japanese. The study is focused on beginner and intermediate learners of Japanese.

More information will be provided to those who wish to join. Please enter your email address, and I will contact you shortly. You can also email me directly at em4733ga-s@student.lu.se or contact me on WhatsApp at +46 723 242 076.

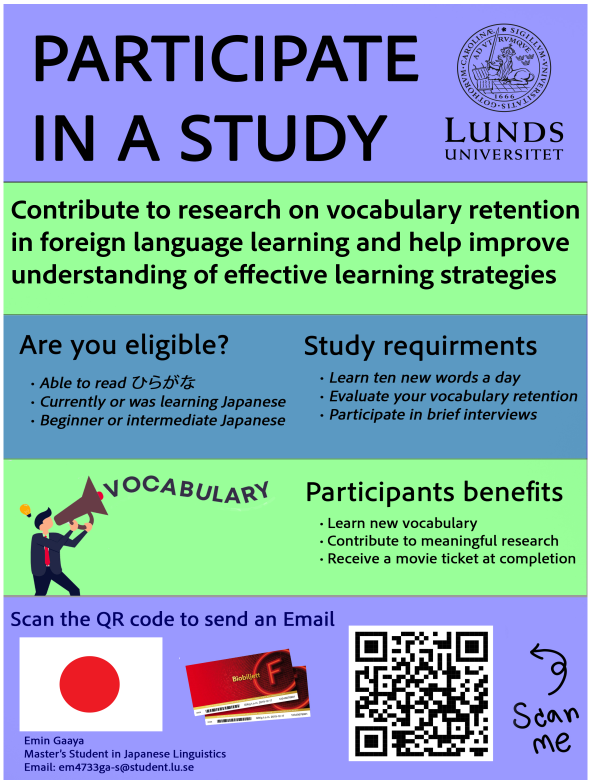
* Indicates required question

Email Adress

Your answer

Interested in joining the study *

○ Yes



PARTICIPATE IN A STUDY

LUNDS UNIVERSITET

Contribute to research on vocabulary retention in foreign language learning and help improve understanding of effective learning strategies

Are you eligible?

- *Able to read ひらがな*
- *Currently or was learning Japanese*
- *Beginner or intermediate Japanese*

Study requirments

- *Learn ten new words a day*
- *Evaluate your vocabulary retention*
- *Participate in brief interviews*

Participants benefits

- Learn new vocabulary
- Contribute to meaningful research
- Receive a movie ticket at completion

Scan the QR code to send an Email

Emin Gaaya
Master's Student in Japanese Linguistics
Email: em4733ga-s@student.lu.se

Figure 4 (Study interest form)

Figure 5 (Flier)

During the recruitment period, the researcher changed the inclusion and exclusion criteria. Initially, we set three inclusion criteria: (a) respondents should have little or no prior experience with Japanese, (b) they must be able to read hiragana, and (c) they should be first-semester university students studying Japanese. We later changed these criteria because they were too restrictive to recruit enough respondents for the study. We broadened criteria (a) and (c) to increase our ability to find more respondents. Since many university students often enter their second semester, we realized that few true beginners could join. We expanded criterion (c) to include second-semester students and eventually adjusted it to include any active or prior learners of Japanese. We also broadened criterion (a) to allow anyone with limited to no experience, regardless of their Japanese level, with eligibility assessed by the pre-test. Criterion (b), the ability to read hiragana, stayed the only inclusion criterion not changed during the study.

Exclusion criteria included respondents who knew all the vocabulary on the pre-test, those who had Japanese as their first language, and anyone unable to meet the inclusion criteria. The current study would have been easier to conduct if students could receive class credit as an award for finishing the study or if teachers had been able to collaborate, as seen in earlier studies (e.g., Yüksel, 2020; Shahipanah et al., 2025; Hanson & Brown, 2019).

About 90 Japanese learners expressed interest in taking part and received a consent form that they could return digitally or in person. The consent form outlined the study's general guidelines and what we expected from them as respondents (Appendix A). We addressed any questions before they signed the consent form. However, we could not provide information about the condition s they would join or the vocabulary items. We could only mention minor details about how we would conduct the study online using Google Form. We also informed the respondents about their right to anonymity, and they could leave the study at any time without any restrictions. Additionally, each respondent needed access to a device (such as a smartphone or laptop) capable of running Anki and commit to studying 10 words daily.

Of the original 90 respondents who expressed interest in taking part in the study, only fifty-three completed the pre-test. The researcher asked three respondents to withdraw due to high pre-test scores. The remaining fifty respondents received an email assigning them to one of two conditions: Control or Anki. During the treatment phase, 10 respondents either stopped communicating or formally withdrew from the study via email. However, the researcher anticipated that some respondents would drop out before finishing the study because of the troubles with online studies (e.g., Hanson & Brown, 2019). Many respondents reported missing daily study sessions and asked whether they should catch up by studying more. The researcher recommended that they resume the planned schedule without catching up. The researcher communicated with respondents once a week during the treatment phase. However, some respondents took several days to reply to first contacts. Future researchers should consider sending reminders through email or SMS (Short Message Service) and explore alternative communication methods to keep respondents engaged and ensure they follow the study procedures effectively with online respondents. In the end, forty respondents completed all tests. However, six respondents could not provide enough data, or there was a long gap between the final study session and the post-test, leading to their exclusion from the analysis.

The final respondents (n = 34) that completed the whole study ranged in age from 20–40 years (M = 26.79, SD = 4.41). They revealed a diverse range of first languages (L1), including Swedish, Norwegian, English, Spanish, German, Luxembourgish, Croatian, Danish, and Russian. On average, respondents had lived in Japan for about 0.37 years (SD = 0.57, range = 0–2.23). Around one-third of the respondents (n = 14) studied Japanese at a university, while 12 had taken classes in the past, and eight were entirely self-taught. The gender distribution included 21 males, 12 females, and one non-binary respondent. The respondents (n = 24) rated themselves between JLPT N4 and N3 levels, while the remaining 10 respondents fell within the range of beginner (N5) to early intermediate (N2). A few respondents (n = 6) also mentioned taking the JLPT, with one of the respondents taking the N2, four taking the N3, and one taking the N5. Most of the respondents came from Sweden (n = 19), followed by Norway (n = 3), Denmark (n = 2), and the

USA (n = 2), with one respondent each from Spain, Luxembourg, Germany, Croatia, China, Canada, and Russia. Many respondents had prior experience with Anki (n = 22), often combining it with other platforms such as Quizlet (n = 12) and Memrise (n = 6). A few also used other tools like WaniKani, Duolingo, or Bunpro, while a minority of respondents reported no experience with any spaced repetition system (SRS) (n = 6).

The Anki condition and control condition included the same number of respondents, with 17 respondents in each condition. The Anki condition used only Anki, while the Control condition could use any vocabulary study strategy except Anki. The researchers provided the Anki condition with brief information through an instruction sheet and a YouTube video showing how to make a new profile on Anki. We answered any other questions via email that the previous information did not provide.

In contrast, the Control condition received an instruction sheet, a vocabulary list, and a study log. Researchers instructed this condition that studying for even five minutes a day was sufficient. The consent form also said that respondents did not need to study for more than 30 minutes daily. Some respondents interpreted this guideline and chose to study for up to 30 minutes a day. Additionally, the Control condition could skip words during their study sessions, while the Anki condition had to review all the words, including those they already knew. However, allowing the control condition to skip known words may be unfair unless the Anki condition had the same option. However, determining when a learner truly knows a word is difficult, and in the Anki condition, words that were answered correctly only a few times were already scheduled for review at intervals extending beyond the four-week treatment.

4.2 Material and instrument

4.2.1 Vocabulary selection process

The study used one hundred and fifty Japanese–English word pairs as target vocabulary. We selected the vocabulary according to various criteria to ensure a useful and representative list for language learners of Japanese. (1) Only vocabulary deemed useful for both intermediate and beginner learners of Japanese could be selected, and words found in three textbooks corresponding to all levels of the Japanese-Language Proficiency Test (JLPT) were deemed useful for the study. The JLPT is a standardized test used in Japan to evaluate the Japanese language proficiency of non-native speakers. The test consists of five levels, with N5 being the most basic and N1 the most difficult. (2) Only nouns were selected, simplifying what respondents needed to learn during the study. (3) The study included only words that respondents could write in hiragana and had kanji characters; it excluded all katakana words and words without kanji characters. (4) Most words had a single translation to minimize ambiguity, though this was not always perfectly achieved.

Applying these four criteria, the initial selection included 500 words of verbs, adjectives, and nouns. Only nouns remained after applying criterion (2), excluding all adjectives and verbs. The remaining vocabulary was 300 nouns, including both hiragana-only and katakana words. Following criterion (3), all hiragana-only and katakana words were excluded from the study. The remaining 150 vocabulary items were then selected from all five levels of the JLPT, as specified in criterion (1). These vocabulary items were divided according to JLPT levels, with adjustments to ensure a balanced list across the diverse levels. Seventy-five words were taken from the third edition of *Genki* (Banno et al., 2020), a textbook commonly used by first-year Japanese students that covers JLPT levels N5 and N4. Fifty words were drawn from *Tobira* (Oka et al., 2009), which intermediate learners of Japanese typically use that covers JLPT N3 and N2. The remaining 25 words were selected from *Shin Kanzen Master: JLPT N1 Preparation* (3A Network, 2011). Together, this set includes terms from all JLPT levels, providing learners with a broad range of useful vocabulary to support their language journey.

4.2.2 Creation of the wordlist

When the word selection process ended, the researcher created a list containing all target vocabulary items used in the study. Each item included its translation and the corresponding Kanji characters. We organized the list adding ten vocabulary items per page to encourage respondents to study 10 new words per day, and to mimic Anki's default settings (20 cards per day). The list order also mirrored the Anki condition order to mimic their experience. However, the respondents in the control condition had no limit on what order they could study the vocabulary, and they could study as many words as they preferred. To check study habits in control condition, respondents recorded their daily study time and the date at the top of the document (see appendix B). We collected these self-reported logs and analyzed as part of the study's data to compare study times between the two conditions with the post-test scores similar to previous studies (e.g., Kornell, 2009; Mondria, 2003; Pyc & Rawson, 2007).

Additionally, we separated orthographic, or semantically related words so they did not appear on the same page in the control conditions word list, nor were they scheduled to be introduced on the same day in the Anki condition. We based the decision on findings by Nakata

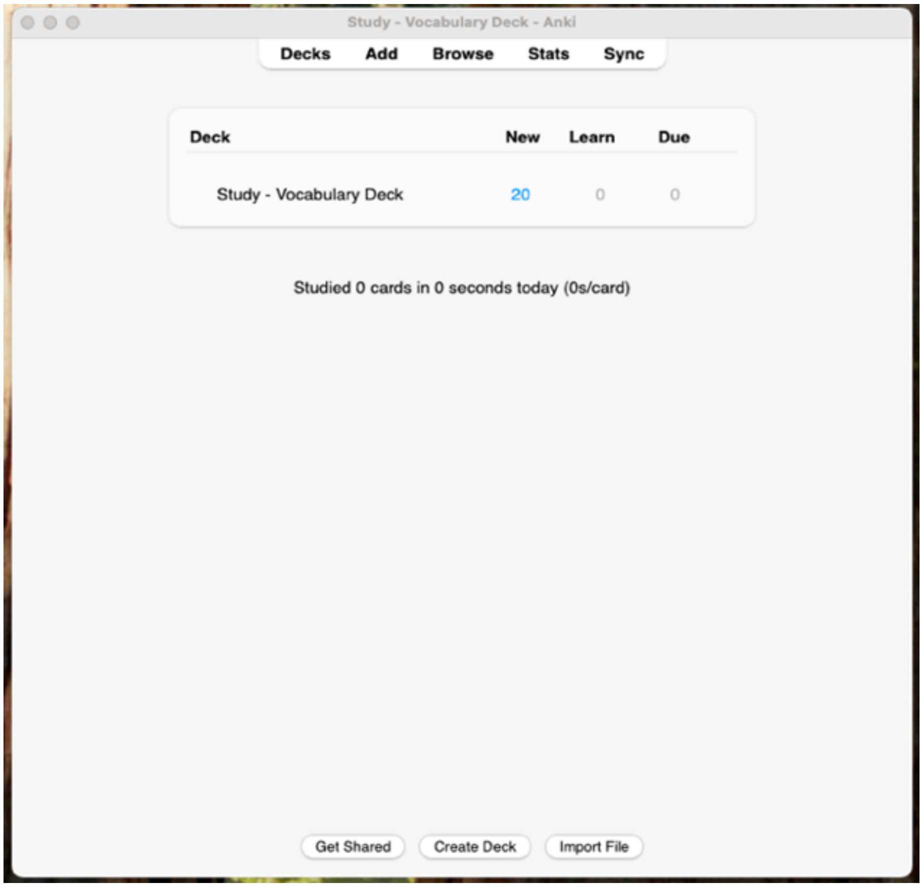
(2019), which suggest that related words can lead to greater interference and hinder retention compared to unrelated words. We intended to hinder interference caused by semantic or orthographic similarity (e.g., でんとう and でんせつ).

4.2.3 Creation of the flashcards in Anki

Anki is an open-source spaced repetition system (SRS) that supports long-term retention through digital flashcards incorporating spaced repetition. The term "Anki" means "memorization" in Japanese, and the developer developed Anki to be a language learning tool, which has expanded and become especially popular among medical students. Anki is available on four platforms: desktop, iOS (AnkiMobile), Android (AnkiDroid), and browser-based (AnkiWeb). All versions provide the same core functionality and allow users to synchronize their study progress across devices using a single user account. Respondents assigned to the Anki condition could use or mix any of these versions in this study.

Creating flashcards in Anki, takes a small amount of time, depending on how much information one wants to add. The addition of add-ons makes the process even easier, creating flashcards automatically. The current study used two flashcards to practice productive and receptive recall. Productive recall flashcards require the respondents to recall the L2 word from the L1 word meaning (Happiness____). Receptive recall flashcards require the respondents to recall the L1 word meaning from the L2 word form (うれしい____). The two flashcards also required the respondents to type the answer, an adaptation from an earlier study (Nakata, 2011) that is not part of the original Anki settings. Instead, we added code to gain this functionality. On the other hand, the respondents did not favour this approach as one needed to switch keyboard language to answer the flashcards. But, according to Nakata, the retrieval effort hypotheses states that adding more retrieval efforts such as typing the answer forces learners to engage with the precise orthographic form, which strengthens form-meaning connections to have the most returns on one's efforts (Pyc & Rawson, 2009; Nakata, 2011).

Anki's interface (see figure 6) is easy to use but has a learning curve for inexperienced users. According to the first interview questionnaire mentioned in section 4.1. Most respondents (n = 22) had used Anki previously, which made teaching them how to use the Anki deck unproblematic. After the researchers randomized the respondents in each condition, and after excluding respondents from the analysis, 11 respondents who had earlier experience with Anki before remained in the Anki condition, and three had used an SRS. Only three respondents had no experience with any SRS but mentioned enjoyment using Anki:" *I generally enjoyed using Anki.*" In the post-interview questionnaire.

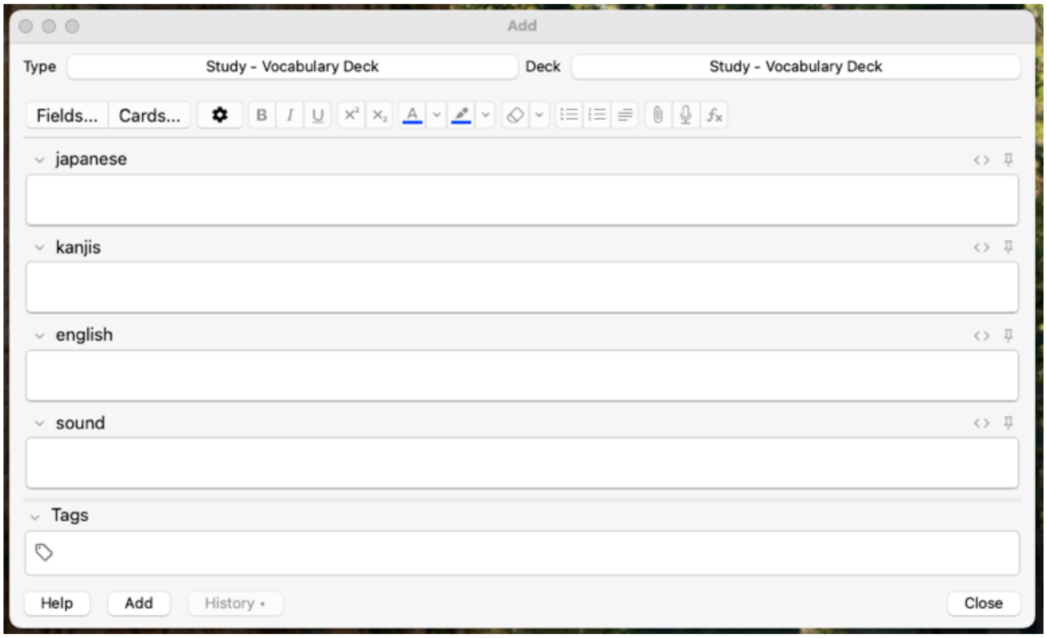


Deck	New	Learn	Due
Study - Vocabulary Deck	20	0	0

Studied 0 cards in 0 seconds today (0s/card)

Figure 6 (Landing page for the Anki desktop (Mac version))

To create flashcards for the study, we started by pressing the 'Add' button on the main interface. After pressing 'Add,' a screen will appear displayed in figure 7. This screen is not the default setting; it holds the fields relevant to our current study. We can rename, add, reposition, and remove the fields by pressing the 'Fields' button, as seen in figure 8.



Type: Study - Vocabulary Deck

Deck: Study - Vocabulary Deck

japanese

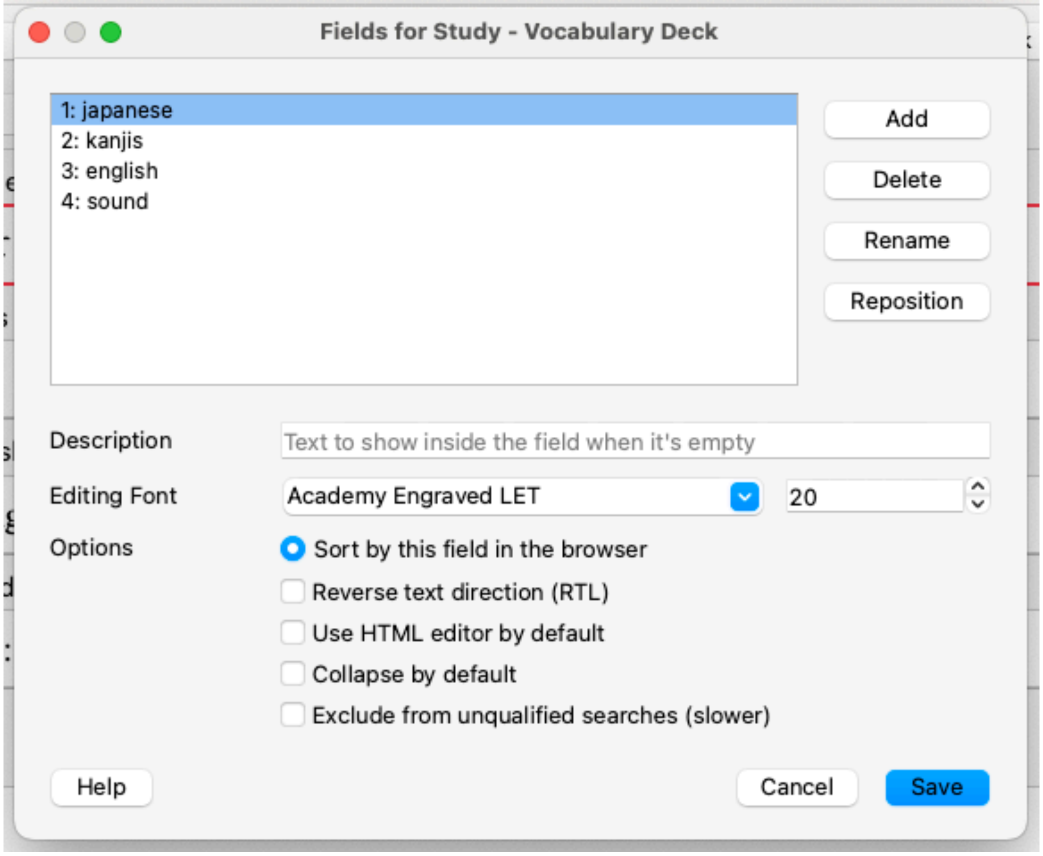
kanjis

english

sound

Tags

Figure 7 (flashcard creation screen)



Fields for Study - Vocabulary Deck

1: japanese
2: kanjis
3: english
4: sound

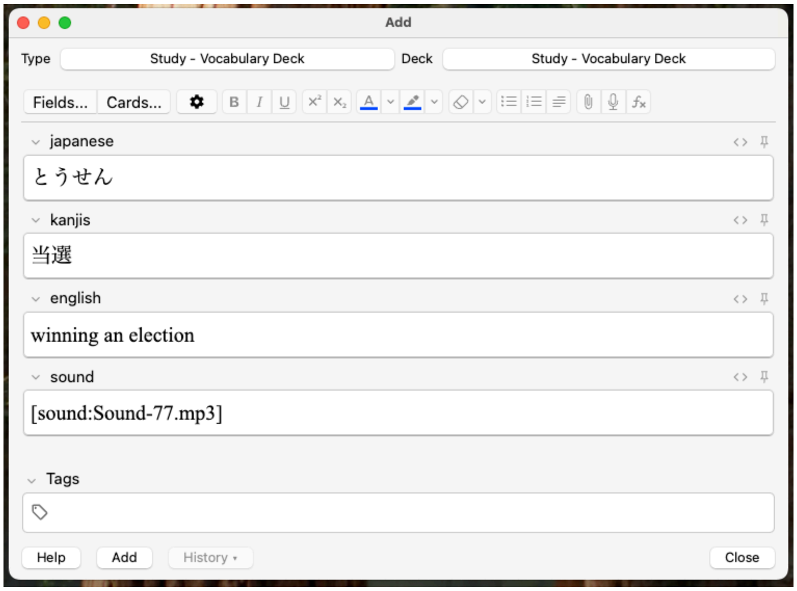
Description: Text to show inside the field when it's empty

Editing Font: Academy Engraved LET 20

Options:
● Sort by this field in the browser
☐ Reverse text direction (RTL)
☐ Use HTML editor by default
☐ Collapse by default
☐ Exclude from unqualified searches (slower)

Figure 8 (Flashcard field editing screen)

For this study, the fields include 'Japanese,' which displays the Japanese word written in Hiragana. 'Kanji,' which displays the Japanese word displayed in Kanji characters. 'English,' which is the translation of the Japanese word, and 'Sound,' which is a recording of a native Japanese speaker from Tokyo pronouncing the word. The researcher recorded the sound using a mobile device in a small room. The recording was 15 minutes, which we cut down to 150 single sound files using the program Audacity. The researcher then using a laptop added the sound into the 'sound' field of the flashcard. Figure 9 displays a completed flashcard with all fields filled out.



Add

Type: Study - Vocabulary Deck

Deck: Study - Vocabulary Deck

japanese: とうせん

kanjis: 当選

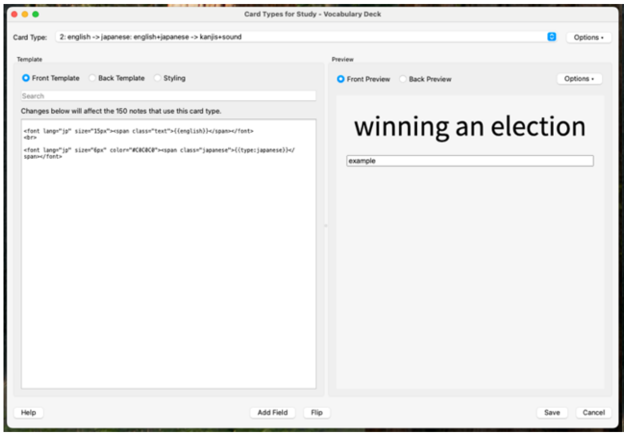
english: winning an election

sound: [sound:Sound-77.mp3]

Tags

Figure 9 (Completed flashcard)

To add the typing feature to the flashcards, you can easily follow the instructions provided in the Anki Manual on Ankiweb. By pressing ‘Cards’ in the flashcard’s creation screen, you can incorporate the necessary code into the flashcards. This code enables respondents to answer the card by typing, as shown in Figure 9 for productive recall and figure 10 for receptive recall.



Card Type: 2: english -> japanese: english+japanese -> kanjis+sound

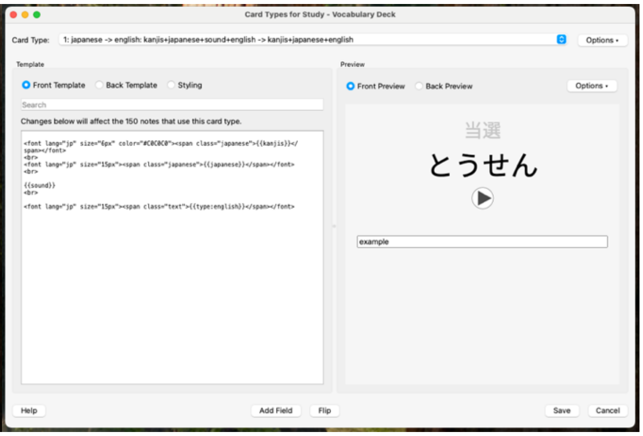
Changes below will affect the 150 notes that use this card type.

```
<font lang="jp" size="15px"><span class="text">{{english}}</span></font>
<br>

<font lang="jp" size="6px" color="#C0C0C0"><span class="japanese">{{type:japanese}}</
span></font>
```

winning an election

example

Figure 10 (productive recall flashcard)



Card Type: 1: japanese -> english: kanjis+japanese+sound+english -> kanjis+japanese+english

Changes below will affect the 150 notes that use this card type.

```
<font lang="jp" size="6px" color="#C0C0C0"><span class="japanese">{{kanjis}}</
span></font>
<br>
<font lang="jp" size="15px"><span class="japanese">{{japanese}}</span></font>
<br>

{{sound}}
<br>

<font lang="jp" size="15px"><span class="text">{{type:english}}</span></font>
```

当選

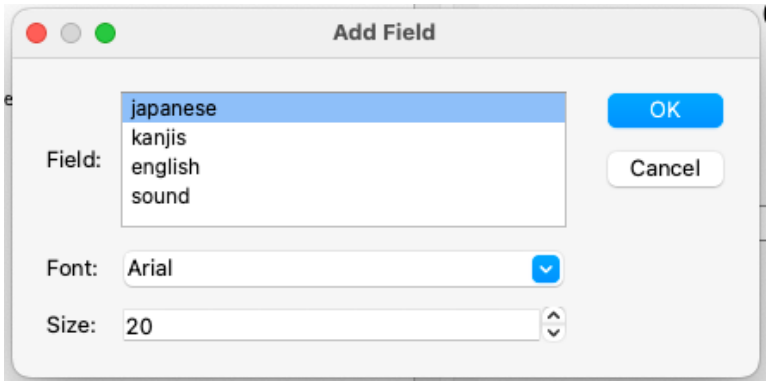
とうせん

example

Figure 11 (Receptive recall flashcard)

Not knowing how to code can intimidate some users, but one can add fields to their flashcards by pressing 'Add Field' in the code editor screen, displayed in Figure 12. One can then add all the fields they want to show on the front of the flashcard. The researcher added 'Kanjis' at the top with half visibility, as we wanted the focus to be on the Hiragana. We then added 'Japanese' and then the 'sound' field. The last field is the word we wanted the respondent to type. We added so the respondents could answer by typing the answer in the flashcard by writing the code 'type:' before the field name, and Anki will do the rest.

<font lang="jp" size="15px"><span class="text">{{**type:**english}}</span></font>



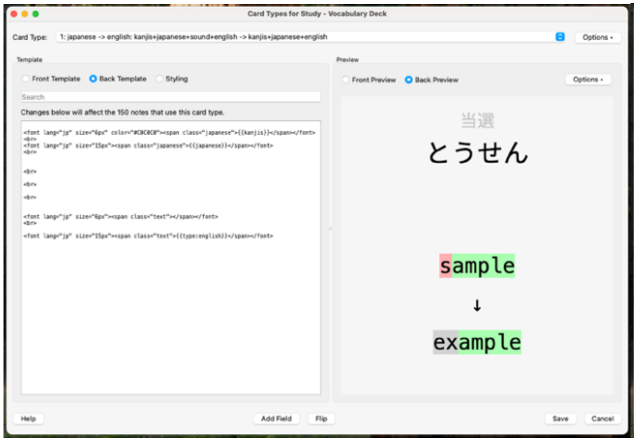
Field: japanese
kanjis
english
sound

Font: Arial

Size: 20

Figure 12 (Add fields to code screen)

We needed to complete the back template to finish the flashcard, as flashcards consist of two sides. To finish the back template, we must add the text {{FrontSide}}, and the flashcard is complete. However, the current study required a different approach for the productive recall flashcard, so there is a distinction between the two codes. We changed the back template to exclude sound since the front template already plays it. Including the sound on the back would be redundant, so we removed it. One can see the back template code in figure 13 for productive recall and figure 14 for receptive recall. Another difference is the position of the Kanji, and we decided that the receptive recall flashcard looked better for it to be under the answer.

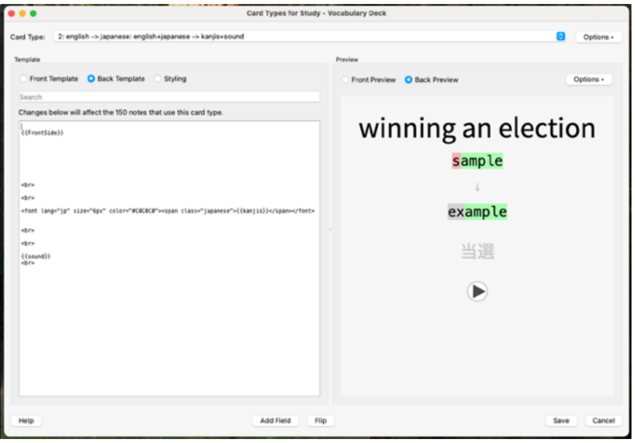

Card Types for Study - Vocabulary Deck
Options ▾
Template
Preview
○ Front Template ● Back Template ○ Styling
○ Front Preview ● Back Preview
Options ▾
Search
当選
とうせん
sample
↓
example
Help
Add Field
Flip
Save
Cancel

Card Type: 1: japanese -> english: kanjis+japanese+sound+english -> kanjis+japanese+english

Changes below will affect the 150 notes that use this card type.

```
<font lang="jp" size="6px" color="#C0C0C0"><span class="japanese">{{kanjis}}</span></font>
<br>
<font lang="jp" size="15px"><span class="japanese">{{japanese}}</span></font>
<br>


<br>

<br>

<br>


<font lang="jp" size="6px"><span class="text"></span></font>
<br>

<font lang="jp" size="15px"><span class="text">{{type:english}}</span></font>
```

Figure 13 (Receptive Recall back layout)


Card Types for Study - Vocabulary Deck
Options ▾
Template
Preview
○ Front Template ● Back Template ○ Styling
○ Front Preview ● Back Preview
Options ▾
Search
winning an election
sample
↓
example
当選
Help
Add Field
Flip
Save
Cancel

Card Type: 2: english -> japanese: english+japanese -> kanjis+sound

Changes below will affect the 150 notes that use this card type.

```
{{FrontSide}}




<br>

<br>

<font lang="jp" size="6px" color="#C0C0C0"><span class="japanese">{{kanjis}}</span></font>


<br>

<br>

{{sound}}
<br>
```

Figure 14 (Productive Recall back layout)

One must create a card type for each flashcard one wants for the vocabulary item. The current study had two card types and thus had two flashcards per vocabulary item with 300 cards. One can add a card type by pressing the 'options' button at the top right of the screen and then pressing the 'add card type' button. We must approve and then done. Figure 15 displays the directions of the two card types. 1: Productive and 2: Receptive recall.

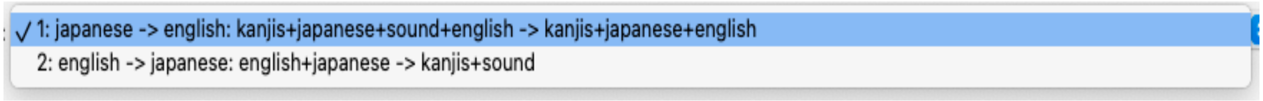

✓ 1: japanese -> english: kanjis+japanese+sound+english -> kanjis+japanese+english
2: english -> japanese: english+japanese -> kanjis+sound

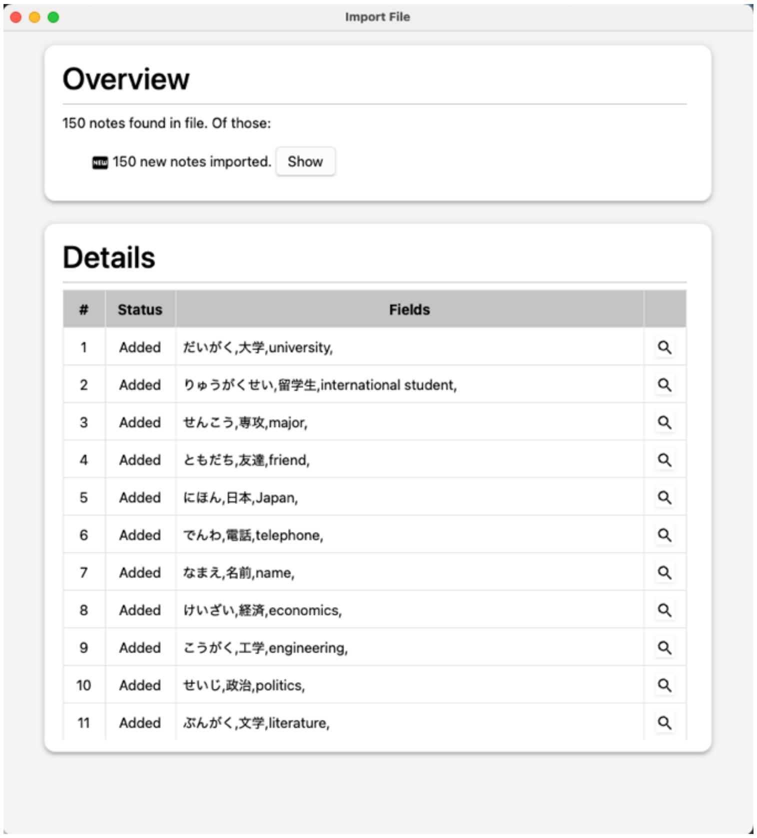
Figure 15 (Add note type menu)

4.2.4 How to use Anki during the treatment

The Anki deck was sent with instructions on how to use the Anki deck before the treatment begun. The respondents could then choose when they started studying the vocabulary. As mentioned previously they received two forms of cards to test productive and receptive recall.

Respondents received instructions on creating a new profile before importing the deck. Current Anki users have specific settings, and creating a new profile resets these settings to the default Anki configuration. By setting this rule, we wanted to minimize the respondents' use of the wrong settings during the study. After creating their profiles, they imported the deck into Anki by either clicking and dragging it directly or double-clicking the Anki deck, which resulted in an automatic import as shown in Figure 16. When the import process succeeded, we can see all 150 new notes as demonstrated in the figure. Each note consisted of two card types for the current study, each standing for one flashcard for a specific word. We learned that one respondent encountered issues and could not import the deck. Currently, we are unsure of the cause. However, the respondents could use the deck on their Android phone via Akidroid, which became their primary device for the study.

When the deck is imported correctly, the default study settings should be applied, which specify 20 new cards per day (10 vocabulary items). However, three respondents reported lower settings of 15, 10, and 5 new cards per day. This occurred because they did not follow the instructions to create a new profile before importing the Anki deck. The respondents received instructions to change setting to the correct 20 new cards per day and then continued studying with the deck. Anki uses distinct colors to indicate the status of each card: new cards (blue) are ones you have not seen before; review cards (green) are those you have seen recently and learning cards (red) are cards seen before but answered incorrectly when trying to recall them.



Import File

Overview

150 notes found in file. Of those:

150 new notes imported. Show

Details

#	Status	Fields	
1	Added	だいがく,大学,university,	🔍
2	Added	りゅうがくせい,留学生,international student,	🔍
3	Added	せんこう,専攻,major,	🔍
4	Added	ともだち,友達,friend,	🔍
5	Added	にほん,日本,Japan,	🔍
6	Added	でんわ,電話,telephone,	🔍
7	Added	なまえ,名前,name,	🔍
8	Added	けいざい,経済,economics,	🔍
9	Added	こうがく,工学,engineering,	🔍
10	Added	せいじ,政治,politics,	🔍
11	Added	ぶんがく,文学,literature,	🔍

Figure 16 (Successful import of Anki deck)

The respondents started using the deck. As mentioned previously, each flashcard came in a fixed order set by the researcher before the study. Productive recall flashcards wanted the respondents to use the hiragana alphabet or click the "show answer" button to receive immediate feedback. The immediate feedback included the L2 word form in hiragana, Kanji characters of the item, a voice recording of the word, and correcting their spelling if they incorrectly typed the answer. Typing an incorrect answer did not automatically penalize the respondent; it only showed the incorrect spelling in red on the feedback screen (see figure 17).

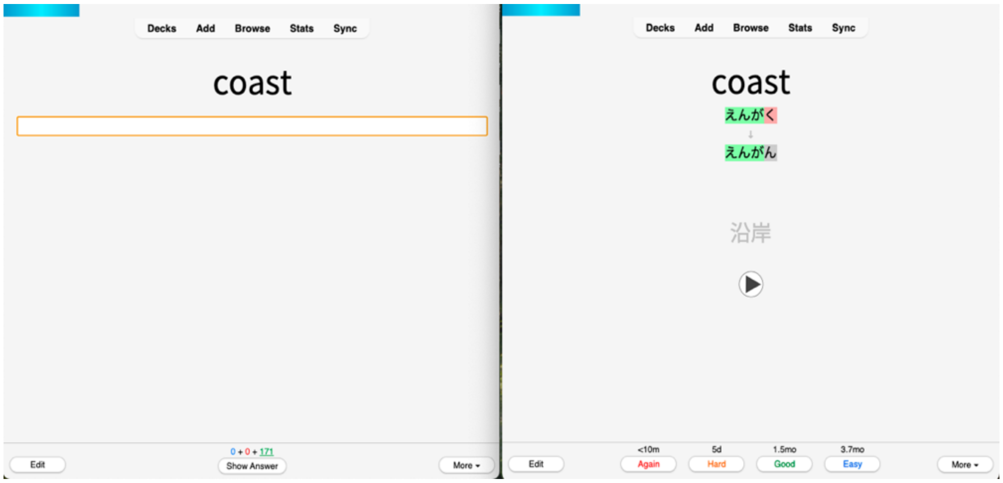


Figure 17 (Productive recall flashcards)

The receptive-recall flashcards displayed the Japanese word in hiragana, Kanji characters, and an audio recording. Respondents then had to produce the L1 meaning of the word by either typing it using the Roman alphabet or clicking the "show answer" button to receive immediate feedback with the L1 translation. We informed the respondents not to be overly concerned about the specific translation in the consent form. However, they answer for the post-test still used the translation from the treatment on the receptive recall test (see Figure 18).

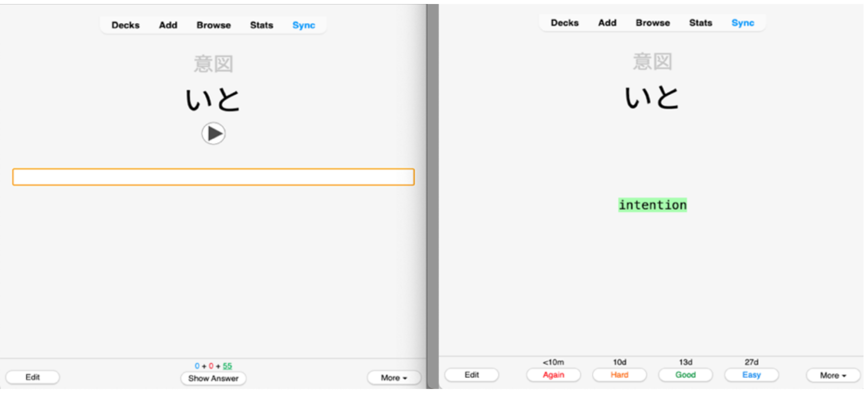


Figure 18 (Receptive recall flashcards)

4.2.5 Anki's algorithm

Anki uses an algorithm based on the SM-2 for *Supermemo*. The algorithm adjusts based on user responses: Each response lets the user self-judge how well they recalled a card. Every card starts with an initial Ease Factor as a percentage (e.g., 250%), which corresponds to a multiplier (2.5) (Vermeer, 2017). The algorithm then adjusted the EF as follows. For pressing the 'Easy' button, the EF increases by 15 percent. For pressing the 'Good' button, the EF stays unchanged, and the interval progresses normally. For pressing 'Hard' button, the EF decreases by 15 percent, and the interval decreases. However, still treated as a correct response. Pressing the 'Again' button, resets the card to a short learning interval for immediate reinforcement (10 min), and the ease factor is decreased by 20%. Each card has its own Ease Factor and its interval, ensuring that the most challenging card for the learner is reviewed more frequently than others.

Anki combines a Leitner-style staging with the SM-2 scheduling once a card is remembered, it '*graduates'* to longer review intervals. The initial step is the learning step of 1 min, 10 min, and 1 day, then it graduates cards, by pushing them to the next day. The card is now at the review stage, and the algorithm kicks in. If a user does not know a card, the algorithm pushes the card back into relearning and stays there until the card graduates from the relearning step again.

An example of a schedule of a difficult card of respondents during the treatment displayed in Table 1. One can also note that the ease factor does not decrease even if the user uses the '*Again*' button more than once during the same session.

Table 1 (Hard Anki card)

Review	Response	EF (Ease Factor)	Next Interval	Day
1	Again	250 %	1 min	March 21
2	Good	250 %	10 min	March 21
3	Good	250 %	1 day	March 21
4	Good	250 %	4 days	March 22
5	Good	250 %	10 days	March 26
6	Again	230 %	10 min	April 5
7	Again	230 %	10 min	April 5
8	Good	230 %	1 day	April 5
9	Good	230 %	3 days	April 6
10	Good	230 %	7 days	April 9

In contrast, with a schedule of an easy card where the respondent does not have difficulty in learning, shown in Table 2. One can also take note that first day the ease factor will not change during the first session even if the user presses the ’Again’ button.

Table 2 (Easy Anki card)

Review	Response	EF (Ease Factor)	Next Interval	Day
1	Again	250 %	1 min	March 15
2	Good	250 %	10 min	March 15
3	Again	250 %	1 min	March 15
4	Good	250 %	10 min	March 15
5	Good	250 %	1 day	March 15
7	Easy	265 %	6 days	March 16
8	Easy	280 %	20 days	March 22
9	Good	280 %	2 months	April 11

The algorithm calculates a flashcard by multiplying the previous interval by the ease factor. Additionally, cards that graduate receive a bonus of 1 day. Anki also applies a bonus that ranges from 1.2 to 1.3, depending on how quickly the user answers the cards. When we see that review three and four of table 1, the interval goes from 1 day to 4 days. The calculations are as follows. 1 day (previous interval) + 1 day (graduation bonus ’learning to review stage’) + (2.5 ease factor × 1.2 bonus) = 4 days. Leaving both 10 min and 1 day leads with a 1-day graduation bonus. However, the algorithm is not that simple and adds an extra factor to cards to minimize sequence effect cards that are always shown together are pushed aside via the algorithm.

4.3 Making an online study

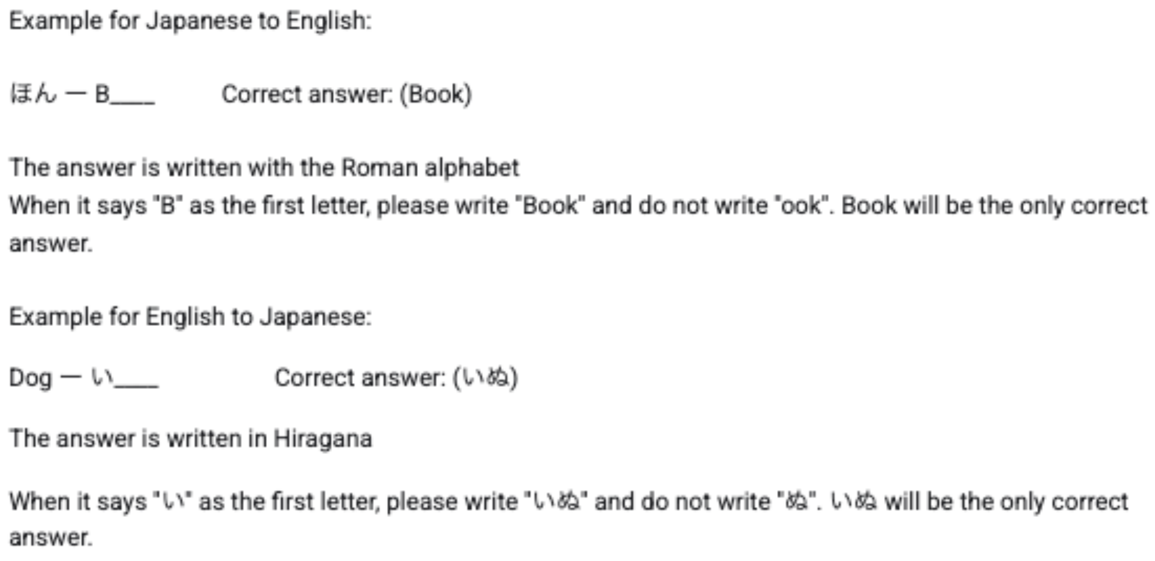
The current study was designed and implemented entirely online to ensure accessibility, as many respondents as possible. This section provides an overview of how the online study environment was constructed, highlighting general considerations that guided its creation. Specific details about instruments and procedures are provided in Sections 4.4 and 4.5.

Choice of platform.

Google Forms was selected as the primary platform because it is free, widely accessible across devices, and requires no installation or specialized training. Alternative platforms were considered, such as PsychoPy, which would have allowed automated scoring through custom code. However, due to the limited timeframe of the project, the need to design and manage three separate tests, and the associated costs of running PsychoPy for each respondent, this option was not feasible. Google Forms therefore offered the most practical balance between, and do have some form of automation of scoring.

Creation of the tests

The study employed five Google Forms: three vocabulary tests (pre-test, post-test, and delayed post-test) and two interviews (pre- and post-). All test items were entered manually, with each question and its corresponding answer key typed into the form builder. To minimize interference errors and potential serial effects, where the order of items can influence memory retrieval with the previous item acting as a cue (Delaney et al., 2011). In addition, each question was presented on its own page. However, this caused that the test could not be be randomized. Instead, each test was randomized once during construction, and that fixed randomized order was delivered uniformly to all respondents. Each test began with standardized instructions on how to complete the task (see Figure 19).



Example for Japanese to English:

ほん ー B____ Correct answer: (Book)

The answer is written with the Roman alphabet
When it says "B" as the first letter, please write "Book" and do not write "ook". Book will be the only correct answer.

Example for English to Japanese:

Dog ー い____ Correct answer: (いぬ)

The answer is written in Hiragana

When it says "い" as the first letter, please write "いぬ" and do not write "ぬ". いぬ will be the only correct answer.

Figure 19 (test instructions)

All vocabulary tests were constructed in the same format but varied in size: 30, 80, and 90 items. In each test, items were split evenly between receptive and productive recall tasks. The pre-test contained 30 items, serving as a baseline. The post-test included 80 items, of which 50 were new vocabulary items not present on the pre-test. The delayed post-test contained 90 items, combining 40 items repeated from the post-test with 50 new items not previously encountered. Across all three tests, respondents were exposed to approximately 120 unique vocabulary items in total, equally divided between receptive and productive formats.

Retrieval Cue

Unique to this study was the use of retrieval cues in the pre-test. Because respondents had no prior knowledge of the specific target vocabulary, cues were incorporated to help them infer the correct answer. This design choice followed Nakata (2011, 2015), who also employed cues in pre-tests to guide recall under similar conditions. In the earlier studies, cues were applied only in the productive recall task (form recall). In the present study, cues were instead included in both the productive and receptive recall tasks. Examples of both the formats can be seen below.

Form recall: The task in form recall is to supply the L2 target word.

"Happiness" - し _ _ _ (target: しあわせ

Meaning recall: The task in meaning recall is to demonstrate a supply the meaning of the L2 word.

しあわせ - H _ _ _ _ _ _ (translation: Happiness)

The current study differed from previous implementations in two important ways. First, because many of our test words have near-identical spellings to synonyms (e.g., both りえき and りじゅん mean "*profit*"), providing the retrieval cue only on the first letter (e.g., "Profit – り_ _ _for りじゅん") did not distinguish between synonyms, unlike in Nakata (2011, 2015). Instead, the retrieval cue should have been on the second letter (e.g., _ じ _ _) to use the retrieval cue effectually. Second, we did not display the total number of letters in each word with underscores and spaces (e.g., り _____), which meant that respondents could not use word length as a supporting cue during recall. These limitations should be considered in future research when designing retrieval cues.

Data handling

All data collection was conducted online, stored securely, and automatically backed up through the university's Google account system. The responses were subsequently exported to a laptop and an iPad for analysis and were deleted after the conclusion of the study.

Study length

The current study lasted for six weeks, consisting of a four-week treatment and a two-week delayed post-test. The duration was chosen to provide both conditions sufficient time to learn the vocabulary and allow for a fair comparison. A total of 150 vocabulary items were selected so that respondents could study ten items per day for 15 days and then review the words requiring the most attention. In the Anki condition, the algorithm guided respondents in determining which words to review, whereas in the control condition, respondents decided on their own study approach.

4.4 Procedure

4.4.1 Setting

The entire study took place online. The respondents did all tests and interviews using Google Forms, and communication with most of the respondents occurred via SMS, WhatsApp, and email. There were no physical meetings with most of the respondents, and some remained completely anonymous, so the researcher does not know their identities. Respondents completed the treatment. at home, during trips, and mostly late at night, reflecting a realistic view of how language learning might occur in everyday life.

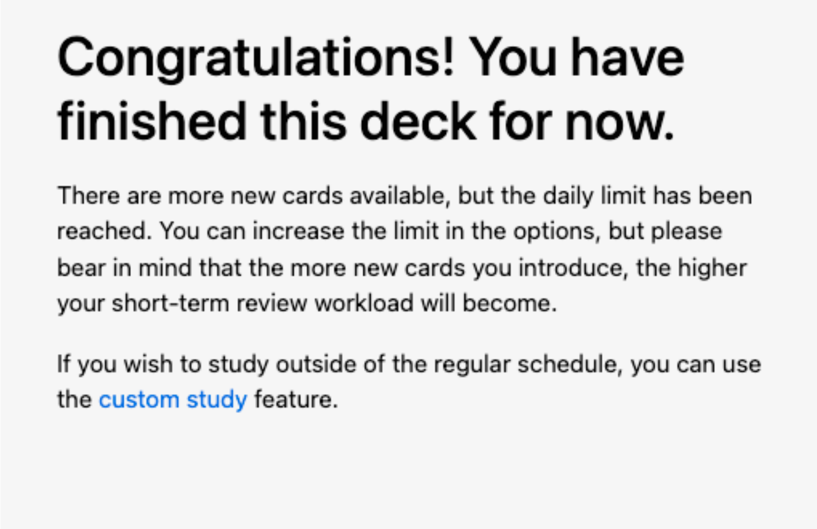
4.4.2 The procedure

Once all the respondents had finished all the pre-study preparations (consent form, questions), each respondent received an email with a Google form link to the study's first phase. The study consisted of a Google Form with a pre-interview questionnaire that automatically redirected them to the pre-test upon completion; there was no time limit for these tasks. Each respondent took their time to finish the task, but the researcher sent a reminder email to anyone who had not completed it. After a respondent completed the pre-interview and pre-test, they advanced to phase two of the study. The researcher randomly assigned each respondent to one of the two conditions, the Anki condition or the control condition. As mentioned previously, each respondent received the vocabulary through Anki or a word list. The researcher informed the respondents to email when they started to study the vocabulary, which indicated the start of the four-week treatment.

Treatment Anki (Phase 2)

The respondents studied 150 Japanese English word pairs using Anki in the Anki condition. The treatment lasted four-weeks, with each respondent working on a single Anki deck and studying 10 new words daily. On the very first day, respondents Anki program introduced the first 10 new words, which amounted to twenty flashcards in total (ten "Japanese-to-English" cards and ten "English-to-Japanese" cards), and studied them until Anki displayed the "Congratulations! You have finished this deck for today" screen (see figure 20). Anki added 10 new words the following day, and the first 10 words were due for review. Respondents began each session by reviewing any cards that Anki's spaced-repetition algorithm had scheduled. If the respondent had correctly remembered a card, Anki would have pushed it to be repeated further into the future. If the respondents had forgotten the card (Pushing the 'Again' button), it would reappear later in that session. The respondents repeated studying the cards until all card finished. The respondents had seen all 150 words after fifteen study sessions. Furthermore, the workload became smaller after each session until only the most difficult vocabulary only remained.

Respondents in the Anki condition were not allowed to study the vocabulary outside the Anki program. Once the daily study session ended. Their study time for that day was considered complete. This restriction did not prevent them from encountering or using the words in natural contexts (e.g., during conversation or reading). However, there were clear instructions from engaging in any active vocabulary study of the 150 vocabulary items outside of Anki. This included writing down the words and using handwritten lists or other materials to review the vocabulary outside of the scheduled Anki sessions. In a more realistic setting, Anki would be combined with some form of short-term retention method, such as massing, to prepare before a test or exam, but the Anki condition did not allow this choice.



Congratulations! You have finished this deck for now.

There are more new cards available, but the daily limit has been reached. You can increase the limit in the options, but please bear in mind that the more new cards you introduce, the higher your short-term review workload will become.

If you wish to study outside of the regular schedule, you can use the custom study feature.

Figure 20 (Anki daily study completion screen)

Treatment control condition (control condition) (Phase 2)

In the control condition, 150 Japanese–English word pairs were studied using a printable PDF or Word document. The treatment lasted four-weeks, with each respondent working from a single word list and studying with no restrictions, except they were not allowed to use Anki. Respondents in the control condition could use any strategy to learn the vocabulary. This included, for example, creating Quizlet flashcards, studying with friends, or writing the words out by hand. Unlike the Anki condition , the control condition followed no tightly controlled parameters or schedules. However, the consent form had instructed each respondent to study at least once daily throughout the four-weeks and document each session using the study log.

Phase three of the study

At the end of the four-week treatment, the researcher contacted each respondent through their contact information and sent two Google form links: one for the post-interview and one for the post-test. Each condition received separate interview questions, asking about their experience. Some questions were condition-specific, while others shared similarities across both condition s. However, the post-test was identical for all respondents, regardless of condition. After the respondents completed the post-interview and post-test, the researcher followed up with any individuals as needed. These follow-ups included clarifications regarding interview responses and questions about the post-test. Once all the respondents had answered the questions, we informed them that they could not study the vocabulary for approximately two weeks. The respondents received no official date for the delayed post-test to keep respondents unaware of the exact timing; while giving them a general sense of the timeframe so they would be available to complete the test when contacted.

Phase four of the study

At the end of the two-week break, the researcher contacted each respondent again via their contact information and sent them two links: one to the delayed post-test and the post interview. After completing the test, the respondents officially finished the study. The researcher thanked the respondents for their assistance over the six-weeks and informed them they would receive a copy of the study once it was complete.

4.5 Data Collection Instrument

To address the research questions, the researcher employed both quantitative and qualitative instruments. Collecting quantitative data using three vocabulary tests administered through Google Forms and gathered qualitative data through two interviews conducted via Google Forms.

Pre-interview questionnaire

The initial screen of the Google Form provided general information, encouraging respondents to answer truthfully, and that the questionnaire included open-ended questions, yes/no questions, and items rated on a five-point Likert scale: strongly disagree, Disagree, I do not know, agree, and Strongly Agree.

The first questions gathered demographic data, including gender, country, and age. The second section asked open-ended questions about the respondent’s language background, including their L1, length of stay in Japan, and how many times they have visited. Additionally, if they are or have been taking Japanese classes, what are their general thoughts about language learning, and if they liked studying Japanese? The section ended by asking them if they had taken the JLPT before and what levels they would rate themselves on.

The third section gathered information about the respondents' study habits, such as how many days a week they study Japanese, what methods they use to study Japanese, and what methods they use in and outside the classroom. Lastly, the section asked if they have ever used an SRS such as *Anki*, *Quizlet*, *Supermemo*, etc.

The final section of the questionnaire is an Attitude/Motivation Test Battery (AMTB) designed to assess the respondents' attitudes and motivation toward learning Japanese (see table 3 for the full question list). The test is a modified version based on three studies, two analyzing English and one analyzing Japanese, adapted from Gardner's (1985) research on motivation and attitudes (Okamura, 1990; Ushida, 2005). The test battery is a five-point Likert scale: strongly disagree, Disagree, I do not know, agree, and Strongly Agree. The questionnaire ended with thanking them and providing a Google Form link to the pre-test.

Japanese-English vocabulary pre-test

The respondents received a link to both the pre-test and the interview. When they finished the interview the completed the pre-test as mentioned in making a study. The data were collected and put in a excel file. Each respondent's score had to be calculated one and one which took days.

Post-interview questionnaire

The two variations of the post-interview questionnaire began by congratulating the respondents for completing the four-week treatment and reminded them to answer all questions truthfully. The questionnaire included a combination of open-ended, yes/no, and multiple-choice questions. The researcher then reminded the respondents that the post-test would follow the questionnaire and that both Google Form links are in the email they have received.

The first section asked all respondents to upload data from their treatment period via the Google Form. The Anki condition exported and uploaded their Anki deck with all study data intact, while the control condition uploaded their study logs. The questionnaire then asked about their general experiences during the four-weeks, for example, if they liked the vocabulary study strategy they used. Additional questions followed, asking when, where, and on what device they studied during the treatment.

This was followed by questions about whether respondents' regular academic work interfered with the treatment. The control condition received additional questions regarding their vocabulary study strategies, such as which words they prioritized during the treatment. The section concluded with questions about whether they would like to integrate their method into classroom settings and whether they considered it an effective tool for studying Japanese.

The second section focused on the use of sound. Since only the Anki condition had sound in their flashcards, the questions are general impressions of sound, such as whether sound helps with retention, whether a native speaker's voice is preferable to AI-generated audio, and whether they found certain Japanese sounds perceptually difficult to distinguish.

The third and final section contained more reflective and specific questions about the respondents' study habits during the treatment. The first question asked how many words they thought they would remember after the upcoming two-week break and whether they had encountered any of those words outside the study context. We also asked whether they believed their Japanese proficiency had improved during the treatment and, if they preferred, a massed learning approach or spaced repetition. The section concluded by asking whether they found the study valuable, and whether the inclusion of Kanji interfered with their ability to learn hiragana. Lastly, the questionnaire ended by thanking the respondents for participating and providing the Google Form link for the post-test.

Japanese-English vocabulary post-test

Following the post-interview questionnaire, the respondents received the post-test. This test is a copy of the pre-test with 50 additional vocabulary items compared to the pre-test, which only had 30. The post-test tested as the pre-test both productive and receptive recall with each form having 40 items each. Respondents completed the post-test on average the day after the treatment concluded (range = 0–3 days, where 0 = same-day completion and 3 = three days later). We expected average accuracy to be around 90% in both condition s. However, the two conditions should differ in their preparation. The control condition was free to review all 150 vocabulary items before testing, whereas the Anki condition could only study those items that Anki had scheduled for review.

Japanese-English vocabulary delayed post-test

The delayed post-test was administered approximately two weeks (15 days) after the post-test. Respondents took on average, one day to complete the test (range: 0–4 days). The delayed post-test included only the most difficult vocabulary items from the treatment materials, comprising 40 words that appeared on the post-test plus 40 additional items, for a total of 90 words.

4.5.1 Scoring of Pre-test and Post-test Protocol

The pre-test and post-test responses were evaluated using two scoring methods: strict and sensitive. These methods were adapted from Nakata's (2013, pp. 50-55), study with some modifications to include scoring of receptive recall and because the languages taught in the original study were opposite (Learners of English instead of learners of Japanese).

In the strict scoring method, only perfectly spelled words in the productive recall test were respondent recalled Japanese target words were considered correct. In contrast, the sensitive scoring method employs what Nakata (2013), calls Lexical Production Scoring Protocol (LPSP, e.g., Barcroft & Rott, 2010; Deconinck et al, 2010), which assigns scores of 0.00, 0.25, 0.50, 0.75, or 1.00 based on how many letters in the response match the target word. ’’A letter is counted as correct if it appears in the exact same position as in the target word’’ (p. 51). However, if another word from the treatment is used as an answer it would be treated as incorrect even if the words are similar.

For the strict scoring of the receptive recall test, misspellings were treated as correct if the intended target word from the treatment was clear (e.g., hygine for hygiene), since the current study is not an English spelling test. Additionally, plural forms or gerund (e.g., speak(ing)) of the target word is treated as correct (p. 145). Finally, only the vocabulary from the treatment is considered as correct and any synonyms are treated as incorrect. In the sensitive scoring method however, synonyms found in a dictionary were also treated as correct and awarded 0.5 points. Allowing learners that already have previous knowledge of the word or remembered a synonym be awarded points for understanding the word. Below is an extraction of the scoring method by Nakata (2013), for the productive recall test with some modifications.

'1.00: all letters in the response are correct. 0.75: 50% or more but less than 100% of the letters in the response are correct. 0.50: 25% or more but less than 50% of the letters in the response are correct. 0.25: at least one letter in the response is correct or 25% or more but less than 0.00: all other responses.’’

Nakata (2013, p. 51)

An example of sensitive LPSP scoring is when a respondent recalls the response けいさい for the target word けいざい (economy). This response receives a score of 0.75 in LPSP because 75% (3 out of 4 letters: けい_い) of the characters in the response are correct. Another example is when a respondent recalls でんとう for the target word でんせつ (folklore). This response is treated as incorrect and does not receive a partial score of 0.50, even though 50% (2 out of 4 letters: でん_ _) of the characters are correct. This is because でんとう is a valid, answer for another vocabulary that is part of the treatment.

An example of sensitive scoring for the receptive recall test is when a respondent produces the response "main character" for しゅじんこう, when the target word is "protagonist." The response receives a score of 0.50, because the respondent used a valid dictionary definition but did not show full recall of the exact target word from the treatment.

4.6 Analysis of the Data

The study used one-way Analysis of Variance (ANOVA) in Excel to determine whether learners using digital flashcards with Anki exhibited differences in vocabulary gains and retention compared to the control condition. Pre- and post-tests were analyzed to assess the total words retained between the conditions. Separate ANOVAs were conducted for gain and retention scores, and effect sizes (η^2) were calculated in Excel to determine statistical significance.

Vocabulary gain was calculated as the difference between post-test performance and pre-test performance. Pre-test scores were adjusted to reflect the overall word distribution in the study (150 words: 50% easy, 35% intermediate, 15% hard). For example, although the pre-test contained 66% easy, 13% intermediate, and 13% hard items, a raw score of 22/30 corresponded to an estimated 62.5% knowledge of the total vocabulary. If a respondent subsequently scored 100% on the post-test, this represented a vocabulary gain of 37.5%. Retention was then calculated as the proportion of this gain still remembered at the delayed post-test. For instance, if the same respondent later scored 88% (91.3% estimated knowledge), the gain was 37.5% (100 − 62.5), the loss was 8.7% (100 − 91.3), and retention was (37.5 − 8.7) ÷ 37.5 × 100 = 76.8%. In other words, the respondent retained approximately 77% of the vocabulary learned during the four-week treatment, corresponding to 28.8% (43 words) of the total study vocabulary. To more easily understand the current study uses this formula to calculate gain, loss and retention.

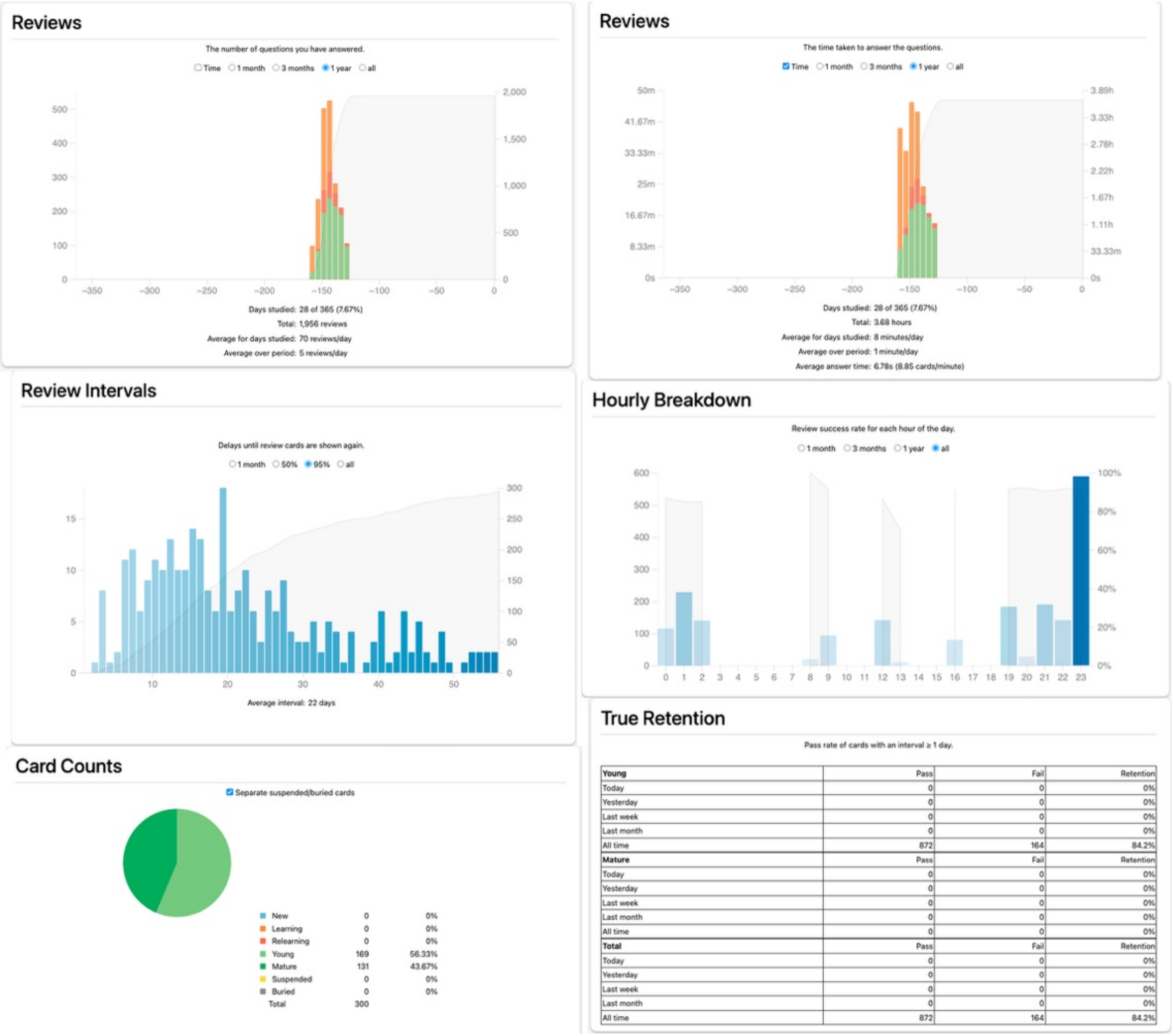
Gain = Post-test − Pre-test = 100% − 62.5% = 37.5%

Loss = Post-test − Delayed Post-test = 100% − 91.3% = 8.7%

Retention = (Gain − Loss) ÷ Gain × 100 = (37.5% − 8.7%) ÷ 37.5% × 100 = 76.8%

The data for the Anki condition were analyzed using Anki's built-in statistics feature. The analysis considered total study time, average daily study time, number of study days, and whether respondents had completed reviewing the entire vocabulary set before the end of the treatment period. As an illustration, figure 21 presents the statistics of one respondent.

In addition, the researcher calculated correlation coefficients in Excel to examine whether individual motivation levels were associated with final vocabulary test scores, specifically testing if more motivated learners achieved higher post-test performance regardless of the study condition. Lastly, semi-structured interviews were conducted to explore respondents' perceptions of Anki's usability and instructional value compared to the control condition.



Reviews

The number of questions you have answered.

☐ Time ○ 1 month ○ 3 months ● 1 year ○ all

Days studied: 28 of 365 (7.67%)
Total: 1,956 reviews
Average for days studied: 70 reviews/day
Average over period: 5 reviews/day

Reviews

The time taken to answer the questions.

☑ Time ○ 1 month ○ 3 months ● 1 year ○ all

Days studied: 28 of 365 (7.67%)
Total: 3.68 hours
Average for days studied: 8 minutes/day
Average over period: 1 minute/day
Average answer time: 6.78s (8.85 cards/minute)

Review Intervals

Delays until review cards are shown again.

○ 1 month ○ 50% ● 95% ○ all

Average interval: 22 days

Hourly Breakdown

Review success rate for each hour of the day.

○ 1 month ○ 3 months ○ 1 year ● all

Card Counts

☑ Separate suspended/buried cards

🟦 New	0	0%
🟧 Learning	0	0%
🟥 Relearning	0	0%
🟩 Young	169	56.33%
🟩 Mature	131	43.67%
🟨 Suspended	0	0%
⬛ Buried	0	0%
Total	300	

True Retention

Pass rate of cards with an interval ≥ 1 day.

Young	Pass	Fail	Retention
Today	0	0	0%
Yesterday	0	0	0%
Last week	0	0	0%
Last month	0	0	0%
All time	872	164	84.2%
Mature	Pass	Fail	Retention
Today	0	0	0%
Yesterday	0	0	0%
Last week	0	0	0%
Last month	0	0	0%
All time	0	0	0%
Total	Pass	Fail	Retention
Today	0	0	0%
Yesterday	0	0	0%
Last week	0	0	0%
Last month	0	0	0%
All time	872	164	84.2%

Figure 21 (Anki Sats)

5. Results and Findings

This chapter presents the results in relation to the three-research question: differences in vocabulary gains, differences in retention, and learners' perceptions of Anki compared to the control condition. The analysis begins with potential influencing factors such as pre-test results, study time, interference errors, and motivation. Only pre-test results had a notable effect. Post-test data show that respondents in the Anki condition achieved an average gain 10 points higher than those in the control condition, addressing the first research question. Delayed post-test results further reveal about 30% higher retention for the Anki condition, answering the second research question. Finally, interview findings indicate a positive experience and perceptions for both the Anki condition and the control condition with a slightly more negative experience for the control condition, addressing the third research question.

5.1 Pre-test results

The pre-test served as a baseline measure for each respondent prior to participation in the study. It comprised 30 items: 15 assessing receptive vocabulary knowledge and 15 assessing productive vocabulary knowledge. Of these, 20 items represented N5 and N4 vocabulary, 5 items represented N3 and N2 vocabulary, and 5 items represented N1 vocabulary. Respondents who answered all 20 N5 and N4 items correctly were considered to know approximately 50% of the vocabulary targeted in the study (mentioned in section 4). While it was theoretically possible for a respondent to know higher-level vocabulary without knowing beginner-level vocabulary, such as knowing a N3 vocabulary but not an N5 vocabulary. This did not happen in the current study.

The pre-test results indicated that respondents in both conditions knew approximately the same amount of vocabulary prior to the study. Using the strict scoring method, the control condition scored an average of 18 points (SD = 4) out of 30, while the Anki condition also scored an average of 18 points (SD = 3.67) out of 30. No difference was found between the strict and sensitive scoring methods on the pre-test. A one-way ANOVA revealed no significant difference in pre-test results between the two conditions, *F*(*p* = .99, $\eta^2 = -.03$).

The results indicate that the respondents knew on average around 45% of the target vocabulary at the start of the study. Therefore, it was expected for respondents to answer at least 40 questions correct on average on the post-test, because of their prior knowledge shown on the pre-test results. Figure 22 displays the distribution of pre-test scores across both conditions. The distribution reveals variation between individuals, with some respondents demonstrating relatively high familiarity with the items and others beginning with more limited knowledge.

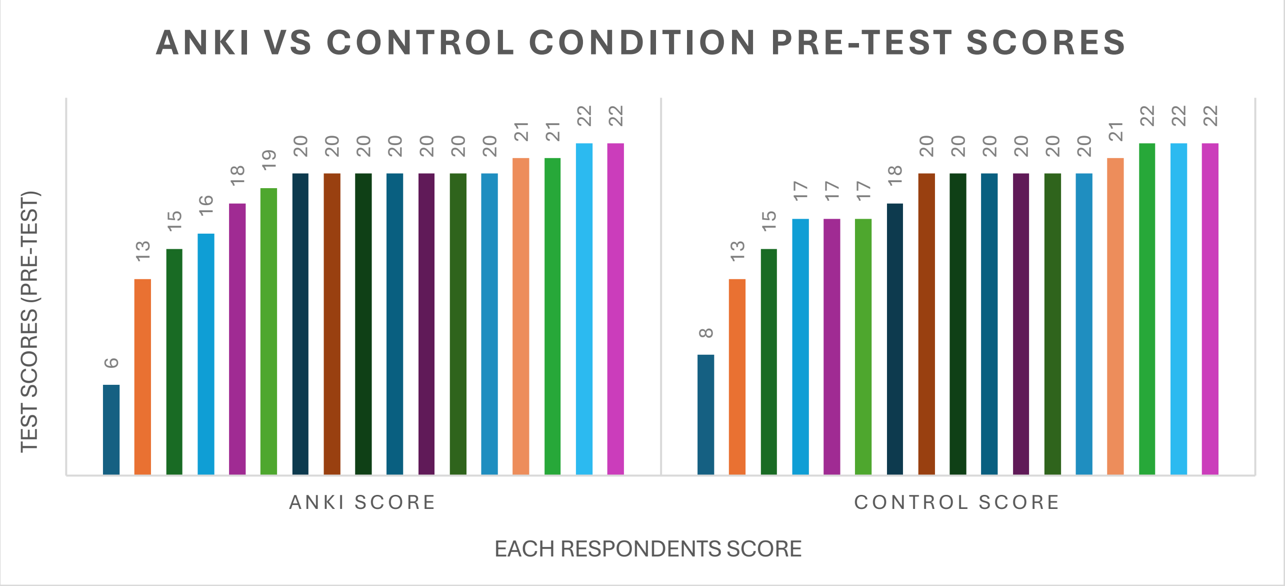


Figure 22 (Pre-test scores distribution)

5.2 Study time

The current study did not impose restrictions on the duration of study sessions, with one exception: respondents in the Anki condition were required to follow the instructions provided by the Anki program. Within each session, however, respondents were free to spend as much time as they wished on individual items. For instance, if the program instructed a respondent to study 20 words on a given day, they could take unlimited time for each word but were not permitted to study more than those 20 words. In contrast, respondents in the control condition determined for themselves both the frequency and duration of their study sessions. However, when looking at the results there were no observed correlation between study time and test score gains in the present study, a finding consistent with previous research (Bower & Rutson-Griffiths, 2016).

The study period lasted a total of four-weeks, and respondents studied for an average of 3.7 weeks (SD = 0.68) during this time. Some respondents missed days, most commonly because they forgot to study. The issue of providing optimal reminders daily instead of weekly which was implemented in the current study. When examining the conditions separately, the data show that the control condition, on average, studied 33% more time than the Anki condition. However, the Anki condition engaged in study on more total days.

The data show that the control condition studied for an average of 23 days (SD = 7.6) at 16.5 minutes per day (SD = 9.5), based on self-reported data, resulting in a total study time of approximately 6 hours and 9 minutes (maximum: 10 hours). In comparison, the Anki condition studied for an average of 27 days (SD = 1.6) at 9 minutes per day (SD = 7.0), according to Anki statistics, for a total of approximately 4 hours and 13 minutes (maximum: 16 hours. Looking at the Anki statistics mentioned in Section 4.6, one respondent in the Anki condition accounted for a disproportionately high total study time. With this respondent included the difference between conditions was not statistically significant at ($p < .05$, $\eta^2 = .10$). Without this respondent, the difference was statistically significant, ($p = .003$). Previous research (e.g., Kornell, 2009; Pyc & Rawson, 2007) has calculated efficiency scores by dividing delayed post-test results by study time in minutes. However, since the results in the present study were not statistically significant, this method was excluded from the analysis.

Looking also into correlation the data showed that both post-test and delayed post-test scores was not statistically significant at ($p < .05$). Figure 23 and 24 illustrates the relationship between study time and scores on the post-test and delayed post-test, indicating that increased study time did not result in significant gains or higher retention. It remains unclear why study time did not have an impact on post-test scores, even within the same condition. Study time data indicated that respondents in the Anki condition typically accumulated around two hours of total study time, suggesting that the Anki condition may have required less time than the control condition to achieve similar results. However, this pattern is not conclusive. In contrast, six respondents in the control condition reported nearly ten hours of study time, yet their test scores varied widely. This may suggest that increased study time reflects weaker initial understanding of the material and thus they need to study more. Further research is needed to determine the optimal duration of study sessions when using an SRS program to maximise learning outcomes.

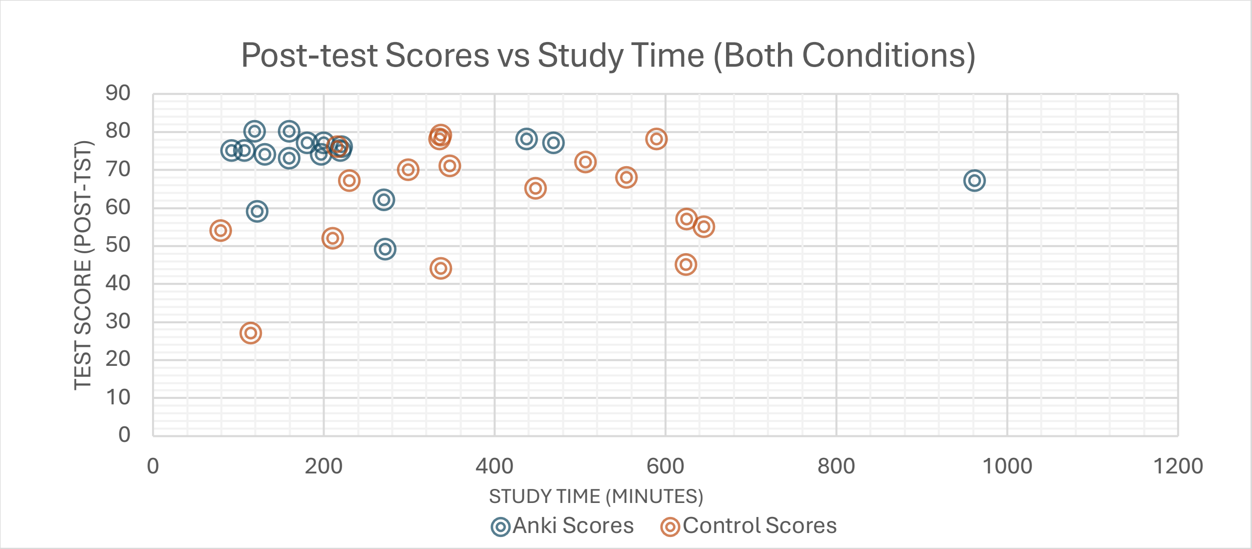


Figure 23 (Study time compared to post-test scores)

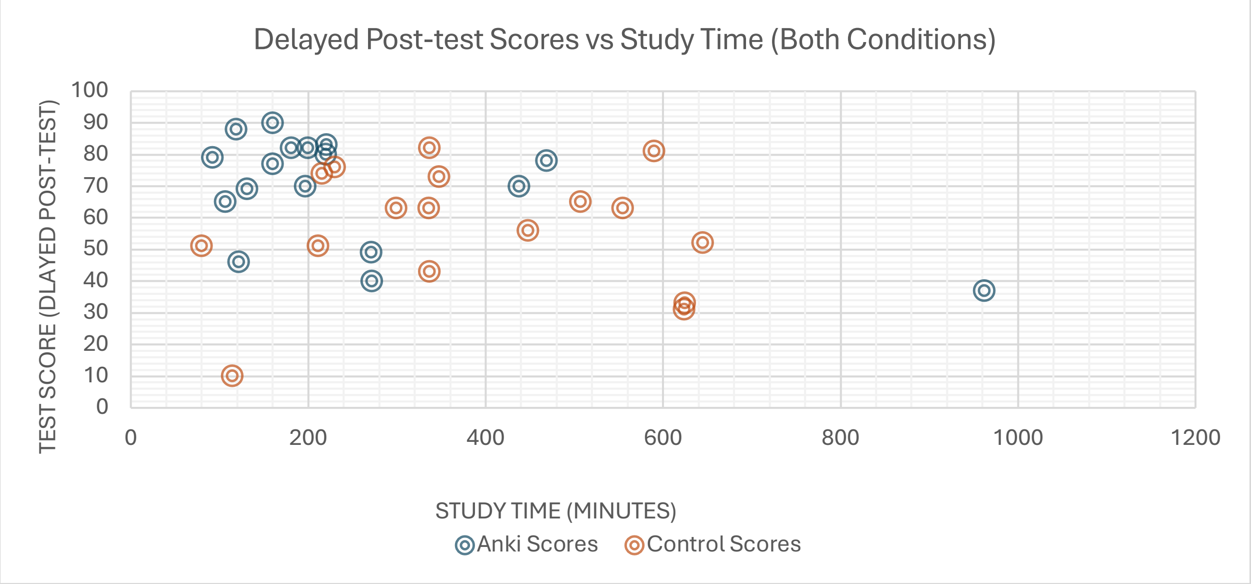


Figure 24 (Study time compared to post-test scores)

As seen in figure 23 and 24, neither the post-test or the delayed post-test showed a significant correlation between study time and test scores. This does not necessarily mean that study time had no influence on post-test performance; rather, the relationship may be influenced by other factors.

5.2 interference errors

Interference errors occur when learners confuse words with similar meanings or forms, often because such words sound alike or share orthographic similarities. This similarity makes them more difficult to learn simultaneously, suggesting that they should be taught separately rather than together (Nation & Webb, 2011; Nakata & Suzuki, 2019). In the present study, related vocabulary was excluded from the word list to minimise this effect. However, despite these efforts, some related pairs still appeared. According to Nakata and Suzuki (2019), semantic clustering occurs when learners are taught words such as *sun* and *moon* together, and subsequently associate the translation for *moon* with *sun*, and the translation for *sun* with *moon*.

In the current study the one of the most frequent types of interference in both conditions were related to similar orthography. Words that looked similar in writing but had different meanings were mixed up. For example, respondents commonly mixed up でんとう (tradition) and でんせつ (folklore), as well as とうひょう (vote) and とうろん (debate). Additionally, Prior knowledge also contributed to errors. In some cases, words that resembled vocabulary previously learned outside the study caused interference. For example, せいさく (policy) was often confused with せいかく (personality), a word learned in beginner Japanese textbooks.

Homonyms introduced another type of interference in the study. The word しゅうかん can mean either “habit” or “week.” While only “habit” was a target vocabulary item in the study, some respondents wrote “week” instead. This is presumed to be for the same reason as before, “week” is a common word found in beginner Japanese textbooks. Kanji is used to easily not mix up homonyms in Japanese, but the current study only tested the vocabulary with Hiragana.

Finally, synonym-related interference occurred in both conditions but was more frequent in the control condition. This typically arose when respondents used a dictionary in addition to the word list to learn the vocabulary and instead remembered a synonym rather than the exact target translation provided in the study. Consequently, some answers were marked incorrect under the strict scoring method but were considered acceptable under the sensitive scoring method. On average, respondents gained one to two additional points when the sensitive scoring method was applied.

Overall, interference errors occurred in both the Anki and control condition. This interference also persisted into the delayed post-test, likely due to the absence of feedback during testing. In other words, respondents often repeated the same incorrect answers, thereby retaining the inaccurate vocabulary. While the overall rate of interference did not differ substantially between conditions, the Anki condition produced more precise responses, suggesting that SRS may support greater accuracy even when interference is present. For instance, with the word とうせん ("winning an election"), respondents in the Anki condition typically produced the complete and precise answer "winning an election," whereas respondents in the control condition gave less precise responses such as "to win an election," "election," "being elected," "elected," or "getting elected." Nakata and Suzuki (2019) note that interference is an inherent aspect of language learning and is typically a short-term problem. However, the scheduling algorithms in SRS programs could be adapted to mitigate such effects by ensuring that semantically or orthographically similar items are introduced separately. Investigating interference errors in SRS-based learning in greater depth could be a fruitful direction for future research.

5.3 Motivation

At the end of the first interview questionnaire, we included a five-point Likert scale survey adapted from Okamura (1990) to assess learners' motivation for studying Japanese. The results are presented in Table 3, with data reported separately for each condition. This questionnaire was introduced to examine whether motivation correlated with higher test scores. For example, it sought to determine whether respondents with higher motivation tended to achieve higher scores.

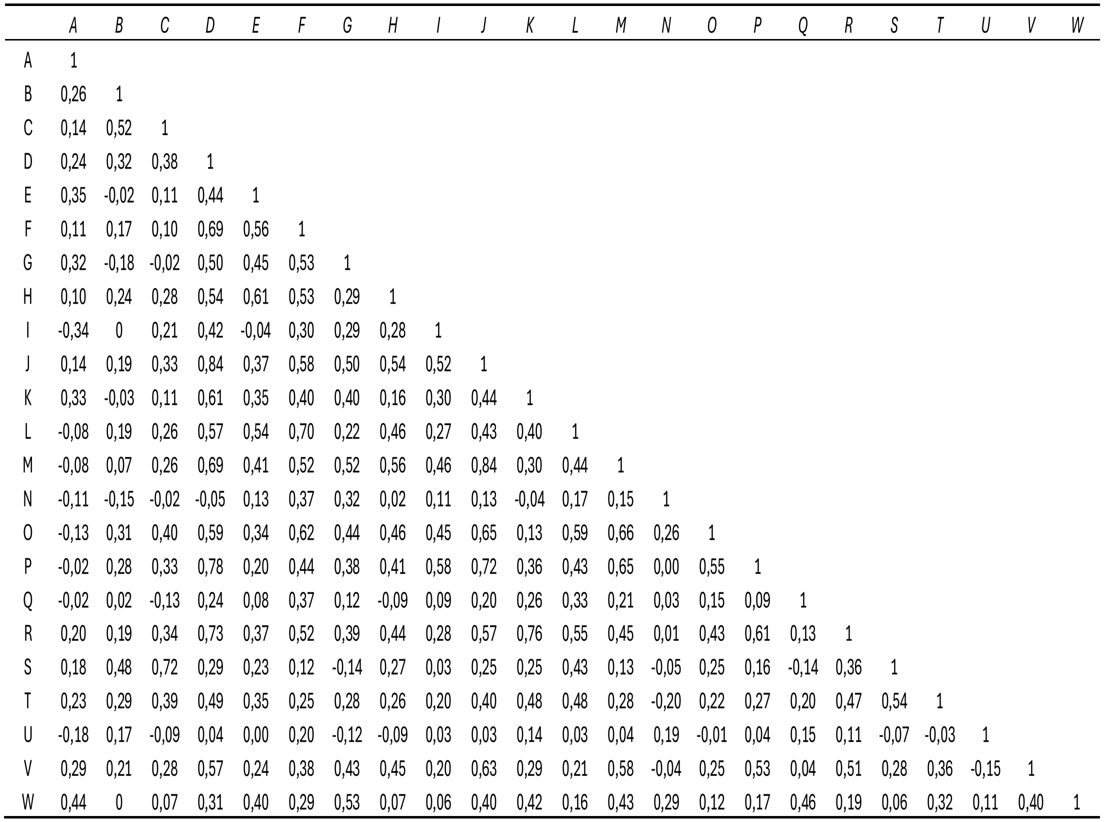
Analysis of the responses indicates that respondents in both conditions strongly agreed that their primary reasons for learning Japanese included (Q) “*to travel to Japan*” and (C) “*interest in the Japanese language*,” which aligns closely with the motivations of Japanese language learners in New Zealand reported in Okamura’s (1990) study. While the two conditions were generally similar in their responses, they differed in the ranking of their motivations. Respondents in the Anki condition most frequently selected (S) “*to read Japanese books, newspapers, or magazines*,” whereas respondents in the control condition most frequently selected (A) “*interest in Japanese culture*.” This was followed, in both conditions, by (T) “*to better understand Japan and its people*,” and finally by (E) “*to connect with various cultures and peoples through Japanese proficiency*.”

In contrast, the reasons for studying Japanese that respondents most frequently disagreed with were (I) “*to get a university degree, and Japanese seemed to be the best way to get one*” in both conditions, (G) “*to contribute to tourism through my job*” in the *Anki* condition, and (V) “*to work for a Japanese company*” in the control condition. This pattern mirrors the findings of Okamura. Respondents in both conditions were generally not interested in studying or working in Japan, but they expressed strong interest in Japanese culture and in travelling to Japan.

Table 3 (Motivation score)

		Anki Condition		Control Condition	
	AMTB Questions	Mean	S	Mean	S
A.	I am interested in Japanese culture.	4.47	0.8	4.76	0.44
B.	I study Japanese because I want to watch animes/dramas/movies in Japanese.	3.88	1.22	4.24	0.75
C.	I am interested in Japanese language.	4.82	0.39	4.88	0.33
D.	I think knowing Japanese will be useful in getting a good job.	2.76	1.48	3.47	1.37
E.	Japanese proficiency is important to me because it will allow me to get to know various cultures and peoples.	3.94	1.39	4.65	0.79
F.	I would like to catch up with, or be with, my friends who are also learning	3.47	1.74	4.0	1.12
G.	I would like to contribute to tourism through my job.	2.12	0.93	2.59	0.94
H.	I would like to help to establish better relations with Japanese people.	3.47	1.28	3.94	0.75
I.	I would like to get a University degree and Japanese seemed to be the best way to get one.	2.06	1.43	1.94	1.03
J.	I would like to get a job which requires Japanese language.	3.0	1.7	2.88	1.22
K.	I study Japanese as much as possible in my free time.	2.24	0.9	2.71	1.05
L.	I would like to improve my communication with Japanese friends or relatives.	3.24	1.6	3.76	1.25
M.	I would like to live in Japan some day.	3.24	1.52	3.29	1.31
N.	I only study Japanese when I have to for class.	1.65	1.11	1.88	0.6
O.	I would like to be able to teach Japanese in the future.	2.76	1.48	2.53	1.18
P.	I'm studying Japanese because it will help me to get a good job.	2.24	1.25	2.47	1.23
Q.	I would like to travel in Japan.	4.94	0.24	4.88	0.33
R.	Learning Japanese is one of the most important things for me right now.	2.94	1.3	3.41	1.12
S.	I would like to be able to read Japanese books, newspapers or magazines.	4.53	0.87	4.65	1.0
T.	I would like to have a better understanding of Japan and the Japanese people.	4.53	0.51	4.76	0.44
U.	No matter how much I study, Japanese is very difficult.	3.71	1.1	3.88	0.78
V.	I would like to work for a Japanese company.	2.47	1.01	2.47	1.01
W.	I'm studying Japanese because I would like to spend a longer period abroad.	3.88	1.27	4.12	1.05
	*On a five point scale ranging from 1 = ''Strongly disagree'' to 5 = ''Strongly agree''.				

The questions also showed patterns of inter-item correlation, meaning that respondents' answers tended to align with other motivation questions. For example, (P) "*I'm studying Japanese because it will help me to get a good job*" and (J) "*I would like to get a job which requires Japanese language*" correlated strongly with (D) "*I think knowing Japanese will be useful in getting a good job*," as respondents seeking Japanese-related employment tended to share these views. Similarly, (F) "*I would like to catch up with or be with my friends who are also learning*" correlated well with (L) "*I would like to improve my communication with Japanese friends or relatives.*" A detailed list of all inter-item correlations, all 23 questions, is presented in figure 25.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
A	1																						
B	0,26	1																					
C	0,14	0,52	1																				
D	0,24	0,32	0,38	1																			
E	0,35	-0,02	0,11	0,44	1																		
F	0,11	0,17	0,10	0,69	0,56	1																	
G	0,32	-0,18	-0,02	0,50	0,45	0,53	1																
H	0,10	0,24	0,28	0,54	0,61	0,53	0,29	1															
I	-0,34	0	0,21	0,42	-0,04	0,30	0,29	0,28	1														
J	0,14	0,19	0,33	0,84	0,37	0,58	0,50	0,54	0,52	1													
K	0,33	-0,03	0,11	0,61	0,35	0,40	0,40	0,16	0,30	0,44	1												
L	-0,08	0,19	0,26	0,57	0,54	0,70	0,22	0,46	0,27	0,43	0,40	1											
M	-0,08	0,07	0,26	0,69	0,41	0,52	0,52	0,56	0,46	0,84	0,30	0,44	1										
N	-0,11	-0,15	-0,02	-0,05	0,13	0,37	0,32	0,02	0,11	0,13	-0,04	0,17	0,15	1									
O	-0,13	0,31	0,40	0,59	0,34	0,62	0,44	0,46	0,45	0,65	0,13	0,59	0,66	0,26	1								
P	-0,02	0,28	0,33	0,78	0,20	0,44	0,38	0,41	0,58	0,72	0,36	0,43	0,65	0,00	0,55	1							
Q	-0,02	0,02	-0,13	0,24	0,08	0,37	0,12	-0,09	0,09	0,20	0,26	0,33	0,21	0,03	0,15	0,09	1						
R	0,20	0,19	0,34	0,73	0,37	0,52	0,39	0,44	0,28	0,57	0,76	0,55	0,45	0,01	0,43	0,61	0,13	1					
S	0,18	0,48	0,72	0,29	0,23	0,12	-0,14	0,27	0,03	0,25	0,25	0,43	0,13	-0,05	0,25	0,16	-0,14	0,36	1				
T	0,23	0,29	0,39	0,49	0,35	0,25	0,28	0,26	0,20	0,40	0,48	0,48	0,28	-0,20	0,22	0,27	0,20	0,47	0,54	1			
U	-0,18	0,17	-0,09	0,04	0,00	0,20	-0,12	-0,09	0,03	0,03	0,14	0,03	0,04	0,19	-0,01	0,04	0,15	0,11	-0,07	-0,03	1		
V	0,29	0,21	0,28	0,57	0,24	0,38	0,43	0,45	0,20	0,63	0,29	0,21	0,58	-0,04	0,25	0,53	0,04	0,51	0,28	0,36	-0,15	1	
W	0,44	0	0,07	0,31	0,40	0,29	0,53	0,07	0,06	0,40	0,42	0,16	0,43	0,29	0,12	0,17	0,46	0,19	0,06	0,32	0,11	0,40	1

Figure 25 (Correlation between the motivation)

In Okamura's (1990) study, motivation was a great indicator of higher test scores. However, in the present study, the *Anki* condition showed a moderate positive correlation, *r*(17 = .41, *p* = .828), which was not statistically significant. In contrast, the control condition showed no correlation, *r*(17 = .06, *p* = .027), which was statistically significant despite motivation had no effect on test scores.

An analysis of raw gains further indicated that respondents with low motivation scores still achieved high test scores. One possible explanation is that not all respondents were current learners of Japanese, unlike in Okamura's study, where all respondents were actively enrolled in Japanese courses. This difference may have contributed to greater variation in motivation, even among those who performed well on the tests. For example, respondents at the top of the motivation scale, with an average score close to four, achieved an average of 68 points on the delayed post-test, whereas those with the lowest motivation scores averaged only one point less, at 67 points. Further, research on motivation and test scores needs to be done for learners of Japanese.

5.4 Post-test results

It was hypothesized in Chapter 3 that the Anki condition would achieve lower post-test scores than the control condition, as spacing effects are generally associated with long-term retention rather than short-term performance. However, contrary to expectations, the Anki condition outperformed the control condition on both the receptive and productive recall tests.

It was also anticipated that respondents scoring around 20 on the pre-test would obtain post-test scores above 40, and this expectation was confirmed. Moreover, results showed a clear trend: the higher the pre-test score, the higher the average post-test performance. Respondents with pre-test scores of 20 or higher in both conditions achieved a mean post-test score of 71 (SD = 7.7) out of 80, compared with 63 (SD = 15.54) for those scoring 20 or below, and approximately 50 out of 80 (SD = 16) for those with pre-test scores of about 15 or lower. Pre-test and post-test scores were positively correlated (r = .37), indicating a moderate relationship between the two measures.

Looking at the post-test scores between the conditions, the Anki condition achieved a mean score of 72 points, compared with about 62 points for the control condition. As there was no significant difference between the conditions on the pre-test, making the difference in post-test performance most likely because of the treatment, rather than the pre-test results. This difference was statistically significant at ($p = .02$, $\eta^2 = .122$), based on the strict scoring method.

Looking also at the sensitive scoring, we see the difference become smaller but still statistically significant as we get at ($p = .04$, $\eta 2$ = ,09). Table 4 show the raw gain scores for both conditions, separated by receptive and productive recall tests. The pre-test had in total 30 questions and the post-test 80 questions divided in two to include the receptive and productive recall test.

Table 4 (productive and receptive results of Pre-test vs Post-test)

Number of Correct Responses Under Both Conditions ***(N = 34)***								
	Receptive				Productive			
	Pretest		Posttest		Pretest		Posttest	
Condition	Correct	%	Correct	%	Correct	%	Correct	%
Anki *(n = 17)*	153	60	611	90	160	62	616	91
Method *(n = 17)*	153	60	526	77	159	62	531	78
Pretest max score = 255 (15 x 17)								
Posttest max score = 680 (40 x 17)								

The pre-test showed that, on average, the respondents knew 45% of the vocabulary prior to the first test. This means they should already know approximately 37 of the 80 vocabulary items on the post-test before even taking it. The post-test represents all vocabulary perfectly to the overall vocabulary ratio in the study, with 50% beginner vocabulary, 35% intermediate vocabulary, and 15% difficult vocabulary. By comparing the pre-test percentage with the post-test percentage, we find that the Anki condition improved by 45%, while the control condition improved by approximately 32%.

5.5 Delayed post-test results

After a two-week period during which no further study was allowed, all respondents completed the delayed post-test to assess vocabulary retention. The control condition achieved on average score of 57 points (SD = 19.4) out of 90, while the Anki condition scored on average of 70 points (SD = 16.7). Table 5 summarizes these results, showing overall scores, in raw gains, with it being statistically significant at (p = .03, η^2 = .103).

Table 5 (Productive and receptive results of post-test vs delayed post-test)

Number of Correct Responses Under Both Conditions ***(N = 34)***								
	Receptive				Productive			
	Posttest		Delayed posttest		Posttest		Delayed posttest	
Condition	Correct	%	Correct	%	Correct	%	Correct	%
Anki *(n = 17)*	611	90	599	78	616	91	587	77
Method *(n = 17)*	526	77	488	63	531	78	478	62
Posttest max score = 680 (40 x 17)								
Delayed posttest max score = 765 (45 x 17)								

Similarly, pre-test scores were a strong predictor of respondents' overall performance on the delayed post-test. Figure 26 illustrates the correlation patterns between pre-test and delayed post-test scores for both conditions. In the Anki condition, the correlation coefficient was strong and positive (*r* = .70, *p* = .001), indicating that higher pre-test scores were associated with higher delayed post-test scores. In contrast, the control condition showed a moderate correlation (r = .45, *p* < .001), suggesting that although pre-test performance still influenced delayed post-test scores, the diversity of study strategies had a greater impact on the results. Looking at retention, the post-test results indicated an average gain of 45% (≈68 words out of 150) for the Anki condition and 32% (≈48 words out of 150) for the control condition. On the delayed post-test, results showed that the Anki condition lost 11% (≈17 words) and retained 34% (≈51 words) of the learned vocabulary, whereas the control condition lost 15% (≈23 words) and retained 17% (≈26 words) on average. Looking also at retention after two weeks, the Anki condition retained 71% of the vocabulary gained (32% retained ÷ 45% gained), whereas the control condition retained 50% of the vocabulary gained (17% retained ÷ 34% gained) on average.

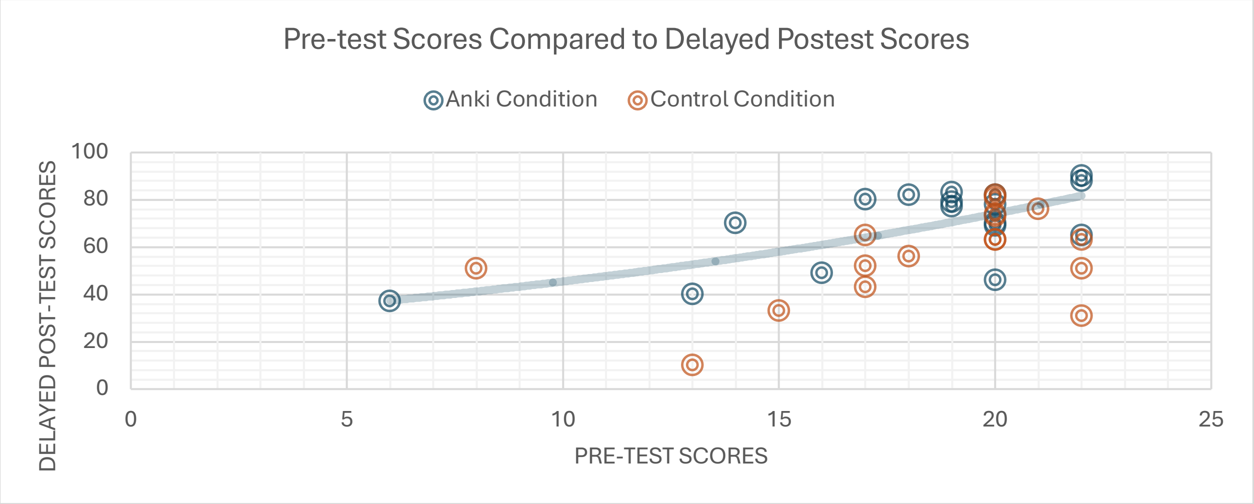


Figure 26 (Pre-test effect on delayed post-test scores)

Ten of the hardest vocabulary

Now that we have examined the overall results, we turn to ten vocabulary items that were included in the pre-test, post-test and the delayed post-test. These items were unfamiliar to the respondents prior to the study, with only a few respondents knowing one or two of them beforehand. In the pre-test, respondents in the Anki condition knew 11 words (6%), while those in the control condition knew 9 words (5%). These ten items comprised five from the productive recall test and five from the receptive recall test.

Looking at raw gains after the four-week study period. The Anki condition learned, on average, 8 out of the 10 new words, after knowing an average of 1 word prior to the tests. In comparison, the control condition learned, on average, 5 of the new vocabulary items, also after knowing around 1 word prior to the tests. If we also compare the recall tests (receptive vs. productive recall), as shown in Table 6, the two conditions performed similarly across tests, with the Anki condition performing slightly better on the receptive recall test than on the productive recall test. A one-way ANOVA indicated the difference between the conditions were statistically significant, ($p = .001$, $\eta^2 = 0.31$).

Table 6 (Pre-test and post-test scores on ten vocabularies)

Number of Correct Responses Under Both Conditions *(N = 34)*								
	Receptive				Productive			
	Pretest		Posttest		Pretest		Posttest	
Condition	Correct	%	Correct	%	Correct	%	Correct	%
Anki *(n = 17)*	4	5	78	91	7	8	71	84
Method *(n = 17)*	4	5	48	56	5	6	47	55
Pretest max score = 85 (5 x 17)								
Posttest max score = 85 (5 x 17)								

Two weeks later the respondents in the Anki condition forgot around 14 (10%) of the new vocabulary and the control condition forgot around 10 (12%) of the new vocabulary. Table 7 shows the difference between the post-test and delayed post-test and the percentage between receptive and the productive recall tests. On a one-way ANOVA indicated that this difference was statistically significant, ($p = .001$, $\eta^2 = .21$).

Table 7 (post-test and delayed post-test on ten vocabularies)

Number of Correct Responses Under Both Conditions ***(N = 34)***								
	Receptive				Productive			
	Posttest		Delayed posttest		Posttest		Delayed posttest	
Condition	Correct	%	Correct	%	Correct	%	Correct	%
Anki *(n = 17)*	78	91	75	88	71	84	61	71
Method *(n = 17)*	48	56	45	53	47	55	41	48
Posttest max score = 85 (5 x 17)								
Delayed posttest max score = 85 (5 x 17)								

The data for these words were calculated by comparing correct responses on the post-test with those on the delayed post-test. Overall, the Anki condition outperformed the control condition on both the recall tests. However, there are some inconsistencies that may have influenced the results. For example, if a respondent already knew a word before the study, its retention cannot be thanks to the treatment, as it was part of their prior knowledge. These words were consistently answered correctly throughout the study. Because they appeared in all test sessions, it was possible to conduct this analysis, something that was not able in the overall analysis of the test data.

Another inconsistency arose when a respondent answered a word incorrectly on the post-test but correctly on the delayed post-test. The two-week period between these tests was intended to be a period without study of the target vocabulary. However, some respondents may have continued studying or encountered the words during their regular studies. To ensure fairness, such cases were counted as "*not retained*," as the improvement could not be confidently attributed to the treatment. The newly calculated data was organized into a table to show the results for the ten vocabulary items.

Table 8 (Post-test and delayed post-test on ten vocabularies with pre-test factor)

Number of Correct Responses Under Both Conditions *(N = 34)*								
	Receptive				Productive			
	Posttest		Delayed posttest		Posttest		Delayed posttest	
Condition	Correct	%	Correct	%	Correct	%	Correct	%
Anki *(n = 17)*	72	84	63	74	61	71	48	56
Method *(n = 17)*	47	55	35	41	39	46	24	28
Posttest max score = 85 (5 x 17)								
Delayed posttest max score = 85 (5 x 17)								

Compared to the previous numbers, both conditions performed slightly worse on the delayed post-test. The Anki condition still performs better on both tests, but the difference between receptive and productive knowledge for both conditions is now much clearer, aligning more closely with previous research that receptive is more easily retained that productive knowledge not shown in the results in section 5.4 and 5.5 (Mondria & Wiersma, 2004 ; Schneider et al., 2002). The gap between the conditions widens slightly, but only by a small margin. On average, the Anki condition scored 78% correct on the post-test, while the control condition scored 50% . In the delayed post-test, the Anki condition dropped from 78% to 65% vocabulary retained, and the control condition from 50% to 37% vocabulary retained. On average, this corresponds to a loss of about one word during the two-week gap, with both the control and Anki conditions showing an average decrease of around 13%. Looking also if the data is significant on a one-way ANOVA show that, ($p = .003$, $\eta^2 = .238$).

In summary, the data indicate the Anki condition resulted in higher vocabulary retention over the two-week period. Furthermore, the results reveal a difference in total gains between the conditions, with the Anki condition achieving at least 30% higher gains than the control condition. These findings and will be examined in greater detail in the following chapter.

5.6 Post interview questionnaire Results

To address the third research question, respondents completed a questionnaire regarding their general experience and perception. Both conditions received similar questions, with slight variations to address condition-specific criteria. For example, because the Anki condition exclusively used Anki, they were not asked which study method they used, whereas this question was mandatory for the control condition. Only one question was presented as a four-point Likert scale ranging from "Yes, it is very enjoyable" to "No, I dislike learning it," while all other questions were open-ended, allowing the respondents to elaborate on their individual experiences, strategies, and challenges.

Research Question 3: What are the respondents' perceptions and experiences regarding the effectiveness of vocabulary learning strategies in the control condition compared to the Anki condition?

5.6.1 Anki condition interview questionnaire results

In the Anki condition, respondents were asked about their perceptions of the program, their overall experience, and their level of enjoyment during the four-week period. The questions, presented in Table 9, focused on perceptions about Anki as a learning tool and its effectiveness for learning Japanese. Each relevant question and answer will be shown in this section to illustrate the respondents' perceptions and experiences.

Table 9 (Perception in the Anki condition)

How did you find using Anki as a vocabulary learning tool?
Do you think Anki is an effective tool for people to learn Japanese?
Would you like to integrate Anki in classes or in parallel to your classes?

The respondents answered very positively about Anki as a vocabulary learning tool. For example, one respondent stated, *"I think it's a good tool for learning vocabulary; it works well."* Another commented, *"It is great! The spaced repetition system (SRS) is very effective."* Similarly, one respondent mentioned, *"I found it very useful and effective! It seems Anki understands how people best remember vocabulary and has designed their app accordingly."* These three quotes reflect the general experience of the whole Anki condition during the four-week study period. The respondents also found Anki convenient to use, noting, *"I like it; I use it every day on my phone, and using it on the computer worked almost as well."*

Finally, an especially important response highlighted both benefits and limitations: *"Anki feels great, and I think I am starting to learn the words better, but I need to use the words outside of Anki to improve my understanding of how they are used."*

The respondents' answers highlighted different aspects of Anki, such as the usefulness of being able to access the program on both their phone and computer, and the view that Anki should be regarded as a supplement to their studies rather than the sole method for studying vocabulary. Most respondents reported studying primarily at home, largely because the Anki deck wanted respondent to type their answer into Anki. Some also studied at work or at a friend's place, using either a PC or a laptop, and then switched to a phone or tablet or vice versa during the four-weeks. The respondents reported studying mostly at night before going to bed, but one did also study in the morning, or just after work.

In addition, the respondents had either studied Japanese in the past or were currently studying the language. While some did not have the goal of actively learning Japanese at the time, all nevertheless made the effort to study during the four-week period. The question of whether Anki is an effective program for language learning, has been discussed in Section 2.4.2. Consistent with earlier responses, Anki was generally viewed as a tool for maintaining vocabulary knowledge and keeping it sharp. As one respondent noted: *"It is a very effective, convenient tool to learn vocab and kanji, so yes. But you will not get fluent using only Anki."* Another respondent added: *"It's an asset for learning Japanese but not as a main source."* Thus, Anki should be regarded as a supplementary tool rather than the primary method of language study, though it may serve as the main resource for vocabulary learning. To be clear over half of the respondent did answer that they thought Anki is an effective tool to learn Japanese. Importantly, more than half of the respondents indicated that they considered Anki to be an effective tool for learning Japanese.

Following this thread, the respondents wanted Anki to be integrated into the classroom, mentioning: *“Yes, I think I would like to do that. Especially with studying the vocabulary we have already been introduced to. And maybe getting better at kanji.”* Another respondent added: *“YES, I wish all teachers handed out anki decks with stuff to memorise, maths, language, geography, history etc. It would do much good for everyone I think.”* However, not all respondents fully agreed with this point, with one commenting: *“No, I think one should use it only if the person wants to.”* Another stated: *“Not that it's a necessicy for the class, but having already finished decks for the material we should learn would be handy.”* The consensus among all respondents except one was that they wanted to integrate Anki into the classroom, or at least having an Anki deck prepared by a teacher, that could be beneficial for everyone studying. However, it should remain optional and never mandatory for student to use this material to study.

The second part of the questionnaire focused on respondents’ experiences, as presented in Table 10. Some of these questions served as follow-ups to the earlier questions on perception.

Table 10 (Experience in the Anki condition)

How did you find your general experience when using Anki?
How did you feel using the Anki deck?
Do you feel you will continue using Anki after this study?
Did you ever feel frustrated using Anki
What did you feel was difficult when using Anki?

Overall, respondents reported positive impressions of Anki, with an mean score of 3.76 out of 4 (SD = 0.43) when asked whether they liked using the program. One respondent commented: *"I really love using it, I just review and then I'm done."* Another noted: *"I was impressed by all of the features. I especially liked that a card could reappear, and that there was a limit to how much you could revise every day."* None of the respondent gave a negative rating on the Likert scale.

Since most respondents had already used Anki prior to the study, there was a consensus that they accepted how the program worked and appreciated its features. As one respondent stated: *"I've used Anki for 3 years, so I'm used to the app. I generally like it, especially when you actually make sure you do the deck every day to not get overloaded.''* However, respondent also acknowledged that Anki requires daily commitment, as reviews can become too many if skipped. This could be a source of stress for some people as one respondent mentioned: *"I have been using Anki for two years now, so it's mostly a daily routine. It can be a source of stress at some times."* This issue can be mitigated by adjusting the number of words studied each day. In Anki, learners can decide freely how many words they wish to review daily. The default setting in Anki is 20 new cards per day, which is the same as the present study. Further, all respondent mentioned that they would continue using Anki after the current study.

The final question asked respondents about frustrations and difficulties they experienced while using the Anki program during the four-week learning period. Respondents did not express negative comments about the program itself, but rather about their own challenges in remembering vocabulary. One respondent noted: *"Looping through the same 5–8 words at the* *same time, failing to remember any of them. Or when it's bedtime but I forgot to do my daily words, so I'd have to stay up a bit longer."* This refers to situations where a word remains in the learning phase and is repeatedly reviewed until the learner recalls it correctly, at which point it is scheduled for the following day.

The most common difficulty mentioned was the need to study every day in order to keep up with the reviews. Another respondent also highlighted interference errors, particularly with similar words beginning with the same hiragana. The respondent stated: *"Again, I'm not sure it has anything to do with the program, but I found it really difficult when words were very similar in Japanese (e.g., starting with the same few hiragana) to remember them or to answer the correct word."* The conclusion will be discussed in the next section, but in general respondent in the Anki condition were satisfied with the program and the Anki deck.

5.6.2 Control condition interview questionnaire results

The control condition, strategies were examined in greater detail to gain a clearer understanding of the methods they employed (see Table 11). This is followed by their overall perceptions (see Table 12) and experiences (see Table 13), in order to address the third research question to be able compare both the conditions.

Table 11 (what method the control condition used)

What method did you use during these 4 weeks?
Is this a method you regularly use, or is it a new method you created for the study?
Did you find it difficult to come up with your own study method, or did you already have one that you regularly use?

The respondents in the control condition employed a variety of methods, with approximately half using some form of physical flashcards to study the vocabulary. Unlike the Anki condition, the control condition required respondents either to design their own study method or to rely on one they had used previously, which may have influenced the consistency of their approaches. Half of the respondents reported creating and using a new method specifically for the study. They described the beginning as somewhat difficult but noted that they adjusted over time. As one respondent stated: *"In the beginning yes. But I think it worked out fine."* The respondents who relied on methods they had used before explained that this was a method they had employed prior to transitioning to Anki: *"I have used it before for practicing flashcards but made the switch to Anki a few years ago due to its convenience."* The control condition may have had hurdles they Anki condition did not, but it was not reported by the respondents.

Other reported methods included the use of a dictionary app, reading and writing, and focusing on the words the respondent did not know yet. The respondent using the app noted: "I added any new and/or difficult words to a dictionary app on my phone called Aedict. I then did daily quizzes on this list, removing any words I felt confident in and adding new words as I went through the list until I completed it. I did two types of quizzes, English to Japanese reading and Kanji to Japanese reading." In contrast another respondent used their pc and mixed the methods to study the vocabulary:''*Created an excel arc on my laptop. Which I created by day 10. Before I just read through the list because I felt I already knew those vocab. /With the excel i read them through top to bottom, day by day. But I also did som flashcard-ish steps: Only looked at the kanji and wrote down the hiragana and translation. To see if i knew them or not. Some i did twice and some once.*''

Some respondents’ method involved some form of the Leitner Box/Leitner System when they studied: *“I used a slightly modified version of a custom SRS method which I found online, called a Leitner Box/Leitner System…* And ‘‘*I wrote each word with on a card with the translation on the backside. Then I went threw the card stack and if I new the translation it got into the next pile. After that I would go threw the next pile and the words I got right would go in a third/finished pile*.’’

The control condition, similar to the Anki condition, studied primarily during the late hours, while a few respondents reported studying in the morning. Most respondents studied at home, although some reported studying while commuting or at their university.

Respondents answered perception-related questions, which are presented in Table 12. These questions were the same as those asked to the Anki condition, and responses are therefore reported in the same format.

Table 12 (Perception in the control condition)

Do you feel about the method you studied the vocabulary is effective? Why or why not?
Do you think your method is an effective tool for people to learn Japanese?
Would you like to integrate your method into classes or in parallel with your classes?

There was a mixed response from the control condition regarding the effectiveness of the methods they chose for vocabulary learning. Respondents who used physical flashcards generally appreciated the method but also noted several practical challenges. As previously mentioned in Section 2.1, physical flashcards can be difficult to keep track of and are often considered bulky and inconvenient. One respondent acknowledged some benefits of physically writing the kanji but emphasized the drawbacks of using physical flashcards, explaining: *"It could be more effective if I didn't constantly rehearse cards that I already know very well… The cards take space… I can't study on the bus or whenever is convenient for me… making the cards takes a non-trivial amount of time."* These limitations highlight several of the advantages of Anki, which eliminates the physical bulk of cards and automatically schedules reviews so that learners avoid repeatedly studying words they already know and more.

Similarly, another respondent commented: *"I feel this method is effective, as it involves creating cards and physically writing words."* However, others who had transitioned to Anki emphasized its greater convenience: *"I use Anki often these days, it's easier to use when you're not home."* Likewise, some respondents who evaluated their methods less favorably expressed concerns about efficiency and motivation: *"I don't think it was very effective, at least in my case. It is very time consuming compared to the method I always use (Anki), and it is harder to know when is the most effective moment to review a word. Also, with this method it is harder to see the progress I'm making and to maintain the motivation,"*

When asked whether they considered their chosen method to be an effective tool for learning Japanese, respondents in the control condition provided mostly positive feedback. One respondent explained*: "It can be effective for some people, but mostly I would say that no, it isn't. It is very tedious, time consuming, difficult to track progress and difficult to maintain the motivation." Another similarly remarked: "Effective compared to some, sure. Effective compared to Anki? No, not really."*

Respondents who relied on physical flashcards acknowledged certain advantages but also highlighted issues of convenience, with one concluding: "It is [effective], but Anki is similar in context and cuts out a lot of the work. I think that creating your own cards and writing the words at least once each is a large benefit." This finding aligns with previous research, as self-created flashcards have been shown to enhance retention more than pre-made flashcards (Dodigovic, 2013; Lei & Reynolds, 2022). The control condition was happy with their methods, but still highlighted problems not reported in the Anki condition.

The control condition provided mixed responses when asked whether they would want to integrate their method into the classroom. Unlike the Anki condition, respondents tended to answer from the perspective of how they would approach vocabulary learning as students rather than the teacher offering their method to learn vocabulary. One respondent explained: "*Yes, I would use this to learn new vocab from classes if I were still a student*." Others emphasized the usefulness of receiving a predetermined list of class-related vocabulary, as one noted: "*Yeah, I guess it's an okay method if the vocabulary is words you come across in class*." However, the majority of respondents did not feel that their chosen method was suitable for integration into the classroom.

The final section of the questionnaire focused on the experience-related questions, which are presented in Table 13. Overall, the methods used in the control condition received a generally a slightly lower evaluation score than the Anki condition, with a mean score of 2.88 out of 4 (SD = 0.78).

Table 13 (Experience in the control condition)

How did you feel about using this method to learn vocabulary?
Do you feel you will continue using your own method after this study?
Did you ever feel frustrated using your own method?

The respondents in the control condition reported their experience as time-consuming and tedious. While the study methods themselves hard to create, the creation of physical flashcards, maintaining an Excel spreadsheet, and keeping track of when to study or which words to repeat proved to be a challenging experience for all respondents. One respondent explained that *“there are some benefits to it, by virtue of involving physical, rather than digital flashcards. However, it is a little bit inconvenient at times since its use is reliant on me being at home, since carrying physical flashcards and a box around is too cumbersome.”* Similarly, another respondent reflected that *“it felt a bit tedious, a bit hard to track my progress and wasn’t as motivating as using Anki, my standard vocabulary learning tool.”* Because of these challenges, several respondents expressed a preference for returning to an app-based method rather than continuing with physical flashcards. One respondent even noted that *“it’s ok, but I would much prefer to use Anki since it’s just more convenient. When I look back at the update that I made to my method I notice that it’s just an attempt at physically remaking Anki.”*

Additionally, another respondent highlighted the inconvenience of studying on the go, explaining: *“A bit annoying, because if I wanted to learn on the way, I would have needed to take the cards in the correct learning order with me. Doing it with Anki, the learning algorithm is there automatically.”* Physical flashcard are a commonly used method, but it is perhaps also the easiest method to upgrade by using one's phone or tablet to either automatically create flashcard with an extension or use a pre-made deck used by another student learning the same language.

Overall, even though some respondents reported challenges, these same individuals often noted that they would continue using their method after completing the study, in combination with other approaches. As one explained: "Yes, but I will also use other methods to help me as well, for example *WaniKani* for kanji." Another similarly reflected: "Yes, it works well enough for me and I find it rather enjoyable." However, not all respondents intended to continue. Some expressed a preference for returning to Anki, with one stating: "I don't think so. I think Anki is the most useful for me at the moment and I will continue with it," while another added: "No. Or maybe sometimes. I like to write it down and check myself. But I think I will let an app check if I am correct or not."

The final question asked respondents about frustrations they experienced during the four-weeks of study. While many of these challenges had already been highlighted in earlier responses, several issues were emphasized again. Respondents noted that it was difficult not having a program that could automatically bring up words they struggled to remember, as one explained: "Yes. It is nice in Anki when the hard cards automatically come up. I missed that." Others pointed to the inconvenience of tracking progress themselves, with one commenting: "Yes, because it was inconvenient and I lost track of my learning progress pretty fast." Finally, one respondent simply described the process as demotivating: "Yes, it was boring, easy to forget to do it, and took too long if I actually wanted to learn all of the 10 words each day." In conclusion, the control condition revealed challenges that were not in the Anki condition; these issues will be discussed further in the following chapter.

6. Discussion and Conclusion

This chapter presents the relevant findings from the previous chapter, followed by recommendations for future research on vocabulary learning with flashcard programs and a reflection on the current limitations of the study. The results showed that, on average, respondent in the Anki condition learned more vocabulary during the treatment and retained more on the delayed post-test compared to the control condition. Furthermore, the interviews revealed that respondents in the Anki condition enjoyed learning vocabulary with Anki more and were more likely to continue using the program.

Discussion

There were two research questions in the study that were answered by the results from the three tests (pre-test, post-test and delayed post-test) administered through Google Form.

1. Is there a significant difference in vocabulary gains between digital flashcard using the spaced repetition software Anki compared to the control condition?
2. Is there a significant difference in vocabulary retention between digital flashcard using the spaced repetition software Anki compared to the control condition?

The first research question was addressed by analyzing the post-test scores, which revealed that respondents in the Anki condition scored higher compared to the control condition. On average, respondents in the Anki condition scored 10 points higher than those in the control condition, and this difference was statistically significant ($p = .02$, $\eta^2 = .122$). Although the treatment affected respondents at varying rates, those in the Anki condition improved by an average of 45% on the post-test, with all respondents scoring above 60%. In comparison, respondents in the control condition improved by an average of 33%, with all scoring above 33% on the post-test. The better performance in the Anki condition may be thanks to the treatment itself; however, it might be other factors other than the treatment that caused the Anki condition to score higher. Overall, the higher post-test performance in the Anki condition cannot be explained by differences in study time, interference errors, or prior knowledge. Rather, it reflects the advantage of having immediate access to a structured, premade learning tool. This finding highlights the pedagogical advantage of offering learners structured resources rather than allowing them to devise their own approaches, particularly when the goal is to maximize vocabulary retention.

The methods used in the control condition varied, but common strategies included the use of physical flashcards and reliance on the word lists through reading and writing the target vocabulary. Previous research has shown that digital flashcard software outperforms both word lists and physical flashcards in terms of vocabulary acquisition and retention (Nakata, 2008). Furthermore, according to the hypothesis, the post-test scores were expected to be higher in the control condition, as spacing does not typically facilitate short-term vocabulary learning. However, previous research has noted that explicit vocabulary learning enables learners to acquire a large amount of vocabulary in a relatively short time (Laufer, 2005; Schmitt, 2008). Which algins with the current study were respondents in the Anki condition retained more of the vocabulary in the four-weeks treatment. In addition, previous research has shown that when spacing is applied effectively, it can facilitate faster and more durable vocabulary acquisition (Steinel et al., 2007; Nakata, 2011). For these reasons, this may explain why the Anki condition still scored better on the post-test than the control condition. Another factor is that the four-week period provided sufficient time not only to study the vocabulary but also to repeatedly review difficult words through the Anki algorithm, which prioritizes vocabulary items that learners found more challenging. Previous research has shown that such features in flashcard programs can support more efficient vocabulary learning, and when combined with statistical feedback, can further enhance learners' motivation (Nakata, 2008; Pyc & Rawson, 2007).

The second research question was examined through the delayed post-test and the analysis of the ten most difficult vocabulary items. Overall, the results indicated that the Anki condition retained 32% of the learned vocabulary, while the control condition retained 17% with a significant (p = .03, η^2 .103), demonstrating that the Anki condition led to greater vocabulary retention than the control condition. In addition, looking at the analysis of the ten most difficult vocabulary items, the Anki condition retained 65%, compared to 37% in the control condition. This difference was highly significant (p = .0034, η^2 = .238). Why these items had a higher rate of retention than the overall results could be because of that they were tested three times, which may have provided additional opportunities for reinforcement, unlike other items that were tested only once or twice. Both these results align with previous research that show, spaced learning leads to stronger long-term retention than non-spaced methods (Nakata, 2008).

Lastly, when examining the recall results, respondents achieved the same scores on both receptive and productive tests. This contrasts with previous research suggesting that receptive vocabulary is generally easier to retain than productive vocabulary (Mondria & Wiersma, 2004; Schneider et al., 2002). However, when focusing specifically on the ten most difficult items that appeared in all tests, receptive vocabulary appeared was retained at a higher rate than productive vocabulary knowledge, as illustrated in Table 8. For the overall results of these difficult items, the difference was statistically significant (p = .0034, η^2 = .238).

The first two research questions showed that the spaced repetition (SRS) program *Anki* had a significant effect on both raw gains and retention of the 150 vocabulary items studied over four-weeks. On average, respondents in the Anki condition gained 68 words and retained 51 of them, losing 17 during the two-week delayed. In comparison, the control condition gained 48 words but lost 23 during the same period. The current study was limited by the number of respondents who participated as many were proficient in Japanese and the study wanted to target beginner Japanese learners of Japanese. In hindsight, the vocabulary selected for the study could have been more difficult, as previous research has shown that respondents often know no of the words before the pre-test (Nakata, 2013). The current study tries to mitigate this limitation by the analysis of the ten most difficult vocabulary items, which respondents did not know prior to the study. These results also showed a statistically significant gain and retention advantage for the Anki condition in this subset. Future research should take greater care when selecting vocabulary items; however, designing a study that reflects more realistic learning conditions is not without value.

The final research question answered by interview questionnaire administrated during the fourth week of the study through Google Form.

3. What are the respondents' perceptions and experiences regarding the effectiveness of vocabulary learning strategies in the control condition compared to the Anki condition?

The two interview questionnaires were analyzed, and the results revealed a wide range of responses regarding the respondents' experiences and their perceptions of their own study strategies. The answers presented in the previous section represent only a small snapshot of the total responses, as it was not possible to include all of them in the current study. Since the interview questions varied somewhat, this led to differences in the responses between the conditions. Nevertheless, all relevant answers provided by the respondents were included in the analysis.

The Anki respondents generally reported a positive perception of the program and, as a result, found studying Japanese more enjoyable during the four-week period. This reaction was consistent among all respondents who used the Anki deck, regardless of whether they were previous users or new to the program. The overall impression was strongly positive, with a mean score of 3.76 out of 4 (SD = 0.43) when respondents were asked whether they liked using the program, and no negative ratings were reported. Several comments reflected this sentiment, such as: *"I really love using it, I just review and then I'm done."* While Anki is widely regarded as a beneficial tool, specifically for vocabulary learning. In contrast, the control condition gave more mixed feedback. While some respondents acknowledged benefits to physical flashcards (e.g., reinforcing memory through writing), they also noted practical challenges such as bulkiness, inefficiency, and the difficulty of tracking progress. The control condition scored slightly lower, with a mean of 2.88 out of 4 (SD = 0.78). Nevertheless, many respondents still saw value in their chosen methods. Physical flashcards were praised for reinforcing learning through handwriting. For example, one respondent noted: *"I think creating your own cards and writing the words at least once each is a large benefit."* This observation aligns with previous research, which has shown that self-created flashcards can enhance vocabulary retention more effectively than pre-made materials (Dodigovic, 2013; Lei & Reynolds, 2022).

In addition, respondents acknowledged certain challenges, most notably the requirement for daily commitment. one noted that reviews could accumulate quickly if skipped, which at times caused stress. As one respondent explained: *"I have been using Anki for two years now, so it's mostly a daily routine. It can be a source of stress at some times."* Others reported difficulties with remembering similar vocabulary items, particularly words beginning with the same hiragana, as well as frustration when repeatedly reviewing the same words until they were successfully recalled. One respondent described this as: *"Looping through the same 5–8 words at the same time, failing to remember any of them."* These issues, however, were attributed more to the inherent demands of vocabulary learning than to flaws in the program itself. Overall, positive impressions of Anki outweighed these minor frustrations, reinforcing its value as a vocabulary learning tool. Control respondents, by contrast, found their self-created methods time-consuming, inconvenient, and often demotivating. They frequently mentioned problems with portability, consistency, and the absence of automatic scheduling. As one respondent reflected: *"When I look back at the update I made to my method I notice that it's just an attempt at physically remaking Anki."*

The findings suggest that Anki and other spaced repetition systems (SRS) can serve as viable and user-friendly tools for educators, with strong potential to enhance vocabulary learning. All respondents in the Anki condition reported finding value in integrating the program into their classroom learning and expressed interest in continuing to use it after the study. They also noted increased motivation, which aligns with previous research showing that new technologies, especially those with multimedia capabilities, can enhance both enjoyment and motivation (Oblinger, 2005; Nakata, 2013). These responses indicate that Anki shows promising results for both educators and learners. Moreover, the simplicity of creating a premade deck containing the required vocabulary can help reduce the workload associated with vocabulary instruction, thereby allowing teachers to devote more time to other aspects of language learning. However, it should be acknowledged that this program may not work equally well for everyone. As some respondents suggested, the use of Anki in the classroom should be viewed as an optimal option rather than a mandatory requirement. In contrasts, the control condition responses were more mixed. While some believed their chosen methods could be useful for studying class-related vocabulary, most expressed doubts about their suitability for classroom integration.

6.1 Limitations of the study

The current study has several limitations that need to be addressed. First, there were some mistakes in the study vocabulary provided to both the Anki and control condition. All respondents were asked whether they were negatively affected by these mistakes, but according to their feedback, they did not notice them. Furthermore, during the pre-test stage, some respondents contacted the researcher to point out errors, which were swiftly corrected. Nevertheless, this meant that some respondents were exposed to mistakes while others were not.

Second, there were major issues related to unclear instructions in both conditions regarding how to complete the treatment. For example, although all Kanji included in the vocabulary list and flashcards were not mandatory to study, this was not clearly stated, and only some respondents received clarification after asking. In addition, the instructions stated that respondents should ignore the definition because a given word could have many translations, yet in practice, the translations provided in the study were treated as the correct answers in the recall tests. This inconsistency may have influenced the way respondents approached the study.

In addition, a lack of consistent communication created further difficulties, such as incorrect settings and skipped days that could have been avoided with clearer guidance. Moreover, because of the online nature of the study, it was not possible to intervene if cheating occurred during the exams, and no measures were in place to prevent this. Some respondents also delayed their work and were slow to reply. For example, the timing of the post-test and delayed post-test varied, with some respondents completing them several days apart duo to respondents' schedules coming in the way. Furthermore, the interview questions were largely unstructured and did not closely align with previous research, which reduced their effectiveness. Nevertheless, open-ended qualitative questions with detailed responses still had the potential to yield valuable insights.

Finally, the lack of respondents who had no prior experience with Anki may have influenced the results, and a more balanced sample could have produced outcomes different from those observed in the present study.

6.2 Recommendations

Based on the results of the study, spaced repetition software (SRS) demonstrates value for vocabulary learning. However, since the current study assessed only receptive and productive vocabulary knowledge through recall tests, future research should employ additional assessment instruments to measure respondents' receptive and productive vocabulary knowledge before and after the treatment.

Second, the current study tested vocabulary learning only through flashcards that presented an L1 definition paired with the target word in the L2. Future research is recommended to incorporate additional information into the flashcards, such as example sentences, images, or video which may help learners encode vocabulary more effectively.

Finally, while the present study demonstrated the spaced repetition system *Anki* to be effective for vocabulary learning, it was limited to nouns. Future research should therefore explore other word classes, such as verbs and adjectives, as well as grammar-related items. In addition, larger sample sizes are needed to increase the generalizability of the findings.

References

Banno, E., Ikeda, Y., Ohno, Y., Shinagawa, C., & Tokashiki, K. (2020). *Genki: An integrated course in elementary Japanese 1* (3rd ed.). The Japan Times.

Barcroft, J., & Rott, S. (2010). Partial word form learning in the written mode in L2 German and Spanish. *Applied Linguistics, 31*(5), 623–650. https://doi.org/10.1093/applin/amq017

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). MIT Press. https://doi.org/10.7551/mitpress/4561.001.0001

Bower, J. V., & Rutson-Griffiths, A. (2016). The relationship between the use of spaced repetition software with a TOEIC word list and TOEIC score gains. *Computer Assisted Language Learning, 29*(7), 1238–1248. https://doi.org/10.1080/09588221.2016.1222444

Bueno-Alastuey, M. C., & Nemeth, K. (2020). Quizlet and podcasts: Effects on vocabulary acquisition. *Computer Assisted Language Learning, 35*(7), 1407–1436. https://doi.org/10.1080/09588221.2020.1802601

Cakmak, G. (2021). Evaluation of scientific quality of YouTube video content related to umbilical hernia. *Cureus, 13*(4), Article e14675. https://doi.org/10.7759/cureus.14675

Carpenter, S. K., & Olson, K. M. (2012). Are pictures good for learning new vocabulary in a foreign language? *Only if you think they are not. Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(1), 92–101. https://doi.org/10.1037/a0024828

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), 354–380. https://doi.org/10.1037/0033-2909.132.3.354

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science, 19*(11), 1095–1102. https://doi.org/10.1111/j.1467-9280.2008.02209.x

Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *Modern Language Journal, 80*(2), 183–198. https://doi.org/10.2307/328635

Cooper, S., Twardowski, N., Vogel, M., Perling, D., & Ryznar, R. (2023). The effect of spaced repetition learning through Anki on medical board exam performance. *International Journal of Medical Students, 11*(4), 271–275. https://doi.org/10.5195/ijms.2023.1549

Deconinck, J., Boers, F., & Eyckmans, J. (2010). Helping learners engage with L2 words: The form-meaning fit. *AILA Review, 23*(1), 95–114. https://doi.org/10.1075/aila.23.06dec

Deng, F., Gluckstein, J. A., & Larsen, D. P. (2015). Student-directed retrieval practice is a predictor of medical licensing examination performance. *Perspectives on Medical Education, 4*(6), 308–313. https://doi.org/10.1007/s40037-015-0220-x

Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. *Psychology of Learning and Motivation, 53*, 63–147. https://doi.org/10.1016/S0079-7421(10)53003-2

Dodigovic, M. (2013). Vocabulary learning with electronic flashcards: Teacher design vs. student design. *Voices in Asia Journal, 1*(1), 15–33.

Dunlosky, J., & O'Brien, A. (2022). The power of successive relearning and how to implement it with fidelity using pencil and paper and web-based programs. *Scholarship of Teaching and Learning in Psychology, 8*(3), 225–235. https://doi.org/10.1037/stl0000233

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans.). Teachers College, Columbia University. (Original work published 1885).

Ellis, N. C. (1995). The psychology of foreign language vocabulary acquisition: Implications for CALL. *Computer Assisted Language Learning, 8*(2–3), 103–128. https://doi.org/10.1080/0958822940080202

Ersoy Özer, E., & Koçoğlu, Z. (2017). Mobil destekli dil öğreniminin kelime öğretiminde kullanımı [The use of mobile-assisted language learning in vocabulary teaching]. *Eurasian Journal of Educational Research, 168(*1), 61–82. https://doi.org/10.1501/Dilder_0000000238

Feng, K., Zhao, X., Liu, J., Cai, Y., Ye, Z., Chen, C., & Xue, G. (2019). Spaced learning enhances episodic memory by increasing neural pattern similarity across repetitions. *The Journal of Neuroscience, 39*(27), 5351–5360. https://doi.org/10.1523/JNEUROSCI.2741-18.2019

Fitzpatrick, T., Al-Qarni, I., & Meara, P. (2008). Intensive vocabulary learning: A case study. *Language Learning Journal, 36*(2), 239–248. https://doi.org/10.1080/09571730802390759

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. Edward Arnold.

Gilbert, M. M., Frommeyer, T. C., Brittain, G. V., Stewart, N. A., Turner, T. M., Stolfi, A., & Parmelee, D. (2023). A cohort study assessing the impact of Anki as a spaced repetition tool on academic performance in medical school. *Medical Science Educator, 33*, 955–962. https://doi.org/10.1007/s40670-023-01826-8

Goldman, M., Bryan, J., & Lucke-Wold, B. (2024). Evidence-based educational algorithm "Anki" for optimization of medical education. *Journal of Biomed Research, 5*(1), 1–7. https://doi.org/10.46439/biomedres.5.037

Gyllstad, H., Sundqvist, P., Sandlund, E., & Källkvist, M. (2023). Effects of word definitions on meaning recall: A multisite intervention in language-diverse second language English classrooms. *Language Learning, 73*(2), 403–444. https://doi.org/10.1111/lang.12527

Harris, D. M., & Chiang, M. (2022). An analysis of Anki usage and strategy of first-year medical students in a structure and function Course. *Cureus, 14*(3), Article e23530. https://doi.org/10.7759/cureus.23530

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal, and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge University Press. https://doi.org/10.1017/CBO9781139524780.011

Iravi, Y., & Malmir, A. (2023). The effect of lexical tools and applications on L2 vocabulary learning: A case of English academic core words. *Innovation in Language Learning and Teaching, 17*(3), 636–649. https://doi.org/10.1080/17501229.2022.2102638

Jape, D., Zhou, J., & Bullock, S. (2022). A spaced-repetition approach to enhance medical student learning and engagement in medical pharmacology. *BMC Medical Education, 22*, Article 337. https://doi.org/10.1186/s12909-022-03324-8

Jia, W., Pack, A., Guan, Y., Zhang, L., & Zou, B. (2023). The influence of game-based learning media on academic English vocabulary learning in the EFL context. *Computer Assisted Language Learning, 38*(5–6), 1341–1365. https://doi.org/10.1080/09588221.2023.2276800

Jaya, Ervan. (2020). *Using Anki (a computer-based flashcard program) in improving student's vocabulary*. [Preprint]. OSF Preprints. https://doi.org/10.31219/osf.io/3j5kc

Kanayama, K. (2020). Is expanding spacing more effective than equal spacing for L2 vocabulary learning? *Annual Review of English Language Education in Japan, 31*, 1–16. https://doi.org/10.20581/arele.31.0_1

Kang, S. H., Lindsey, R. V., Mozer, M. C., & Pashler, H. (2014). Retrieval practice over the long term: Should spacing be expanding or equal-interval? Psychonomic Bulletin & Review, 21(6), 1544–1550. https://doi.org/10.3758/s13423-014-0636-z

Kaitsu, T., & Nakata, T. (2025). Analysis of smartphone-based flashcard apps for second language vocabulary acquisition. *Computer Assisted Language Learning*, 1–31. https://doi.org/10.1080/09588221.2025.2481396

Khoshsima, H., & Khosravi, M. (2021). Vocabulary Retention of EFL Learners through the Application of ANKI, WhatsApp and Traditional Method. *Journal of Foreign Language Teaching and Translation Studies, 6*(4), 77–98. https://doi.org/10.22034/efl.2022.325424.1136

Koleini, N., Boroughani, T., Eslami, Z. R., & Xodabande, I. (2024). Exploring the impacts of mobile-assisted learning on university students' technical vocabulary knowledge. *International Journal of Educational Research Open, 7*. https://doi.org/10.1016/j.ijedro.2024.100344

Kornell, N. (2009). Optimizing learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology, 23*(9), 1297–1317. https://doi.org/10.1002/acp.1537

Kim, S. K., & Webb, S. (2022). The effects of spaced practice on second language learning: A meta-analysis. *Language Learning, 72*(1), 269–319. https://doi.org/10.1111/lang.12479

Lado, R. (1964). *Language teaching: A scientific approach.* McGraw Hill.

Larchen Costuchen, A., Darling, S., & Uytman, C. (2020). Augmented reality and visuospatial bootstrapping for second-language vocabulary recall. *Innovation in Language Learning and Teaching, 15*(4), 352–363. https://doi.org/10.1080/17501229.2020.1806848

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). Academic Press.

Laufer, B. (2005). Focus on form in second language vocabulary learning. *EUROSLA Yearbook, 5*, 223–250.

Laufer, B., & Hulstijn, J. H. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*, 1–26.

Lei, Y., & Reynolds, B. L. (2022). Learning English vocabulary from word cards: A - research synthesis. *Frontiers in Psychology, 13*, Article 984211. https://doi.org/10.3389/psyg.2022.984211

Leitner, S. (1972). *So lernt man lernen [How to learn to learn].* Herder

Levy, J., Ely, K., Lagasca, G., Kausar, H., Patel, D., Andersen, S., Georges, C., & Simanton, E. (2023). Exploring Anki usage among first-year medical students during an anatomy & physiology course: A pilot study. *Journal of Medical Education and Curricular Development, 10.* https://doi.org/10.1177/23821205231205389

Logan, J. M., & Balota, D. A. (2008). Expanded vs. equal interval spaced retrieval practice: Exploring different schedules of spacing and retention interval in younger and older adults. *Aging, Neuropsychology, and Cognition, 15*(3), 257–280. https://doi.org/10.1080/13825580701322171

Lu, M., Farhat, J. H., & Beck Dallaghan, G. L. (2021). Enhanced learning and retention of medical knowledge using the mobile flashcard application Anki. *Medical Science Educator, 31*, 1975–1981. https://doi.org/10.1007/s40670-021-01386-9

McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology, 16*(2), 192–201. https://doi.org/10.1016/0361-476X(91)90037-L

Mondria, J.-A. (2003). The effects of Inferring, verifying, and memorizing on the retention of L2 word meanings: An experimental comparison of the ‘‘meaning-inferred method’’ and the “meaning-given method.” *Studies in Second Language Acquisition*, *25*(4), 473–499. https://doi.org/10.1017/S0272263103000202

Mondria, J. A., & Mondria-de Vries, S. (1994). Efficiently memorizing words with the help of word cards and ‘‘hand computer’’: Theory and applications. *System, 22*(1), 47–57. https://doi.org/10.1016/0346-251x(94)90039-6

Mondria, J. A., & Wiersma, B. (2004). Receptive, productive, and receptive + productive L2 vocabulary learning: What difference does it make? In P. Bogaards, & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition and testing* (pp. 79–100). John Benjamins Publishers. https://doi.org/10.1075/lllt.10.08mon

Mujahidah, Hasanah, N., Yusuf, M., Zulfah, & Fatmasyamsiar, A. A. (2024). The implementation of AnkiApp to improve students’ vocabulary mastery. *Southeast Asia Language Teaching and Learning, 7*(1), 9–18. https://doi.org/10.35307/saltel.v7i1.115

Murre, J. M., & Dros, J. (2015). Replication and Analysis of Ebbinghaus' Forgetting Curve. *PLOS ONE, 10*(7), Article e0120644. https://doi.org/10.1371/journal.pone.0120644

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*, 59–82. https://www.lextutor.ca/cover/papers/nation_2006.pdf

Nakata, T. (2008). English vocabulary learning with word lists, word cards and computers: Implications from cognitive psychology research for optimal spaced learning. *ReCALL, 20*(1), 3–20. https://doi.org/10.1017/S0958344008000219

Nakata, T. (2011). Computer-assisted second language vocabulary learning in a paired-associate paradigm: A critical investigation of flashcard software. *Computer Assisted Language Learning, 24*(1), 17–38. https://doi.org/10.1080/09588221.2010.520675

Nakata, T. (2013). *Optimising second language vocabulary learning from flashcards*. [Unpublished doctoral dissertation]. University of Wellington.

Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning: Does gradually increasing spacing increase vocabulary learning? *Studies in Second Language Acquisition, 37*(4), 677–711 https://doi.org/10.1017/S0272263114000825

Nakata, T. (2019). Learning words with flash cards and word cards. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 304–329). Routledge.

Nakata, T., & Suzuki, Y. (2019). Effects of massing and spacing on the learning of semantically related and unrelated words. *Studies in Second Language Acquisition, 41*(2), 287–311. https://doi.org/10.1017/S0272263118000219

Nakata, T., Suzuki, Y., & He, X. (2023), Costs and benefits of spacing for second language vocabulary learning: Does relearning override the positive and negative effects of spacing? *Language Learning, 73*(3), 799–834. https://doi.org/10.1111/lang.12553

Nender, A. J., Muntuuntu, M., & Rombepajung, P. (2022). Increasing students' vocabulary by using Anki-flashcard. *Journal of Teaching English, Linguistics, and Literature, 1*(6), 707–719.

Oblinger, D. (2005). Learners, learning, and technology. *EDUCAUSE Review, 40*(5), 67–75.

Okamura, Y. (1990). *Motivation and attitudes in learning Japanese in New Zealand* [Master's thesis, University of Canterbury]. UC Library. https://doi.org/10.26021/9819

Oka, M., Lawrence, N., Iwasaki, Y., Kondo, M., & Siegel, M. (2009). *Tobira: Gateway to advanced Japanese: Learning through content and multimedia*. Kuroshio Publishers.

Paivio, A., & Desrochers, A. (1980). A dual-coding approach to bilingual memory. *Canadian Journal of Psychology, 34*(4), 388–399. https://doi.org/10.1037/h0081101

Pimsleur, P. (1967). A memory schedule. *The Modern Language Journal, 51*(2), 73–75. https://doi.org/10.1111/j.1540-4781.1967.tb06700.x

Polly, D., Reinke, L. T., Colonnese, M. W., & Blackwelder, A. (2025). Examining differences between games and pictorial flashcards on multiplication basic fact fluency. *The Journal of Educational Research, 118*(2), 77–89. https://doi.org/10.1080/00220671.2024.2446889

Plass, J. L., & Jones, L. C. (2005). Multimedia Learning in second language acquisition. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 467–488). Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.030

Pyc, M. A., & Rawson, K. A. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory & Cognition, 35*, 1917–1927. https://doi.org/10.3758/BF03192925

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60*(4), 437–447. https://doi.org/10.1016/j.jml.2009.01.004

Ramonda, K. (2022). A double-edged sword: Metaphor and metonymy through pictures for learning idioms. International Review of Applied Linguistics in Language Teaching, 60(3), 523–561. https://doi.org/10.1515/iral-2018-0336

Rogers, J., & Cheung, A. (2020). Input spacing and the learning of L2 vocabulary in a classroom context. *Language Teaching Research, 24*(5), 616–641. https://doi.org/10.1177/1362168818805251

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329–363. https://doi.org/10.1177/1362168808089921

Schmitt, N. (2023). Norbert Schmitt's essential bookshelf: Formulaic language. *Language Teaching, 56*(3), 420–431. https://doi.org/10.1017/S0261444822000039

Schneider, V. I., Healy, A. F., & Bourne, L. E. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *Journal of Memory and Language, 46*(2), 419–440. https://doi.org/10.1006/jmla.2001.2813

Schuetze, U., & Weimer-Stuckmann, G. (2010). Virtual vocabulary: Research and learning in lexical processing. *CALICO Journal, 27*(3), 517–528. https://utppublishing.com/doi/10.11139/cj.27.3.517-528

Schuetze, U., & Weimer-Stuckmann, G. (2011). Retention in SLA Processing. *CALICO Journal, 28*, 460–472. https://utppublishing.com/doi/abs/10.11139/cj.28.2.460-472

Seibert Hanson, A. E., & Brown, C. M. (2019). Enhancing L2 learning through a mobile assisted spaced-repetition tool: An effective but bitter pill? *Computer Assisted Language Learning, 33*(1–2), 133–155. https://doi.org/10.1080/09588221.2018.1552975

Serrano, R., & Huang, H.-Y. (2018). Learning vocabulary through assisted repeated reading: How much time should there be between repetitions of the same text? *TESOL Quarterly, 52*(4), 971–994. https://doi.org/10.1002/tesq.445

Shahipanah, A., Khajavy, G. H., & Elahi Shirvan, M. (2025). The effect of textual and textual-pictorial glosses on incidental vocabulary learning in mobile-assisted listening. *ReCALL*, *37*(1), 79–95. https://doi.org/10.1017/S0958344024000193

Sonbul, S., Macis, M., & Gyllstad, H. (2024). The effect of equal versus expanding spacing practice on the deliberate learning of L2 collocations. TESOL Quarterly. Advance online publication. https://doi.org/10.1002/tesq.3364

Steinel, M. P., Hulstijn, J. H., & Steinel, W. (2007). Second language idiom learning in a paired-associate paradigm: Effects of direction of learning, direction of testing, idiom imageability, and idiom transparency. *Studies in Second Language Acquisition, 29*, 449–484. https://doi.org/10.1017/S0272263107070271

Storm, B. C., Bjork, R. A., & Storm, J. C. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition, 38*(2), 244–253. https://doi.org/10.3758/MC.38.2.244

Strauss, E. J., Markus, D. H., Kingery, M. T., Zuckerman, J., & Egol, K. A. (2019). Orthopaedic Resident Burnout Is Associated with Poor In-Training Examination Performance. *The Journal of Bone and Joint Surgery, 101*(19), Article e102. https://doi.org/10.2106/JBJS.18.00979

Teng, M. F. (2022). The effectiveness of multimedia input on vocabulary learning and retention. *Innovation in Language Learning and Teaching, 17*(3), 738–754. https://doi.org/10.1080/17501229.2022.2131791

Thorndike, E. L. (1908). Memory for paired associates. *Psychological Review, 15*, 122–138. https://doi.org/10.1037/h0073570

Udom, G. (2023). Effects of Leitner's Learning Box (LLB) on Enhancing the Teaching and Learning of The Verbs 'Been and Being' In Primary Schools in the FCT, Nigeria. *International Journal of Research and Innovation in Social Science, 7*(1), 1057–1075. https://doi.org/10.47772/IJRISS.2023.7012080

Ushida, E. (2005). The role of students' attitudes and motivation in second language learning in online language courses. *CALICO Journal, 23*(1), 49–78. https://www.jstor.org/stable/24156232

Vermeer, A. (2017). *Anki essentials: The complete guide to remembering anything with Anki.* Foggy Mountain Pass.

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., & Ozsoy, B. (2008). Distributed rereading can hurt the spacing effect in text memory. *Applied Cognitive Psychology, 22*(5), 685–695. https://doi.org/10.1002/acp.1388

Webb, S., Yanagisawa, A., & Uchihara, T. (2020). How effective are intentional vocabulary-learning activities? A meta-analysis. *Modern Language Journal, 104*(4), 715–738. https://doi.org/10.1111/modl.12671

Vlach, H. A., Sandhofer, C. M., & Bjork, R. A. (2014). Equal spacing and expanding schedules in children's categorization and generalization. *Journal of Experimental Child Psychology, 123*, 129–137. https://doi.org/10.1016/j.jecp.2014.01.004

Whitmer, D. E., Johnson, C. I., & Marraffino, M. D. (2022). Examining two adaptive sequencing approaches for flashcard learning: The tradeoff between training efficiency and long-term retention. In R. A. Sottilare & J. Schwarz (Eds.), *Adaptive instructional systems. Lecture notes in computer science* (pp. 126–139). Springer. https://doi.org/10.1007/978-3-031-05887-5_10

Yamagata, S., Nakata, T., & Rogers, J. (2023). Effects of distributed practice on the acquisition of verb-noun collocations. *Studies in Second Language Acquisition, 45*(2), 291–317. https://doi.org/10.1017/S0272263122000225

Yanagisawa, A. (2016). The effects of receptive and productive word retrieval practice on second language vocabulary learning. *The Journal of the Chubu English Language Education Society, 30*, 139–152. https://doi.org/10.20806/katejournal.30.0_139

Yan, T., & Zhou, D. (2023). The influence of the spacing effect on L2 vocabulary learning: A study on Chinese university students. *System, 115*, Article 103049. https://doi.org/10.1016/j.system.2023.103049

Yüksel, H. G., Mercanoğlu, H. G., & Yılmaz, M. B. (2020). Digital flashcards vs. wordlists for learning technical vocabulary. *Computer Assisted Language Learning, 35*(8), 2001–2017. https://doi.org/10.1080/09588221.2020.1854312

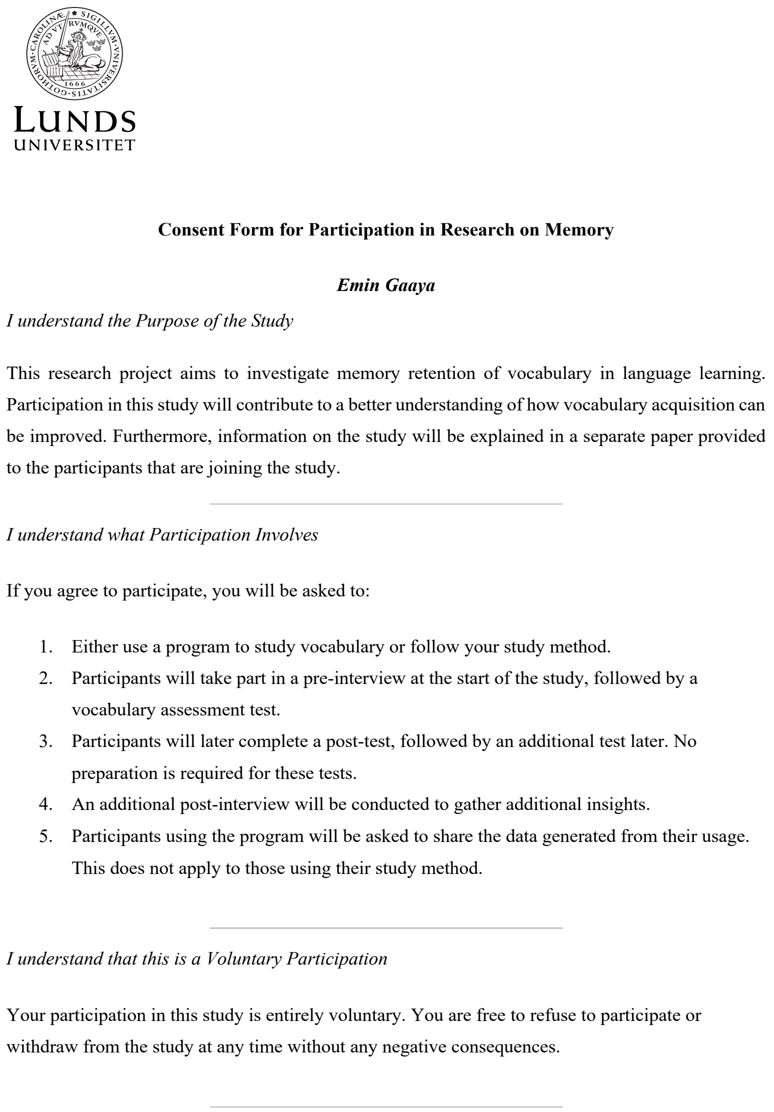
Zung, I., Imundo, M. N., & Pan, S. C. (2022). How do college students use digital flashcards during self-regulated learning? *Memory, 30*(8), 923–941. https://doi.org/10.1080/09658211.2022.2058553

Zulkiply, N. (2013). Effect of interleaving exemplars presented as auditory text on long-term retention in inductive learning. *Procedia - Social and Behavioural Sciences, 97*, 238–245. https://doi.org/10.1016/j.sbspro.2013.10.228

3A Network. (2011). *Shin kanzen master: JLPT N3 vocabulary goi*. 3A Corporation.

Appendices

Appendix A Consent Form



LUNDS
UNIVERSITET

Consent Form for Participation in Research on Memory

Emin Gaaya

I understand the Purpose of the Study

This research project aims to investigate memory retention of vocabulary in language learning. Participation in this study will contribute to a better understanding of how vocabulary acquisition can be improved. Furthermore, information on the study will be explained in a separate paper provided to the participants that are joining the study.

I understand what Participation Involves

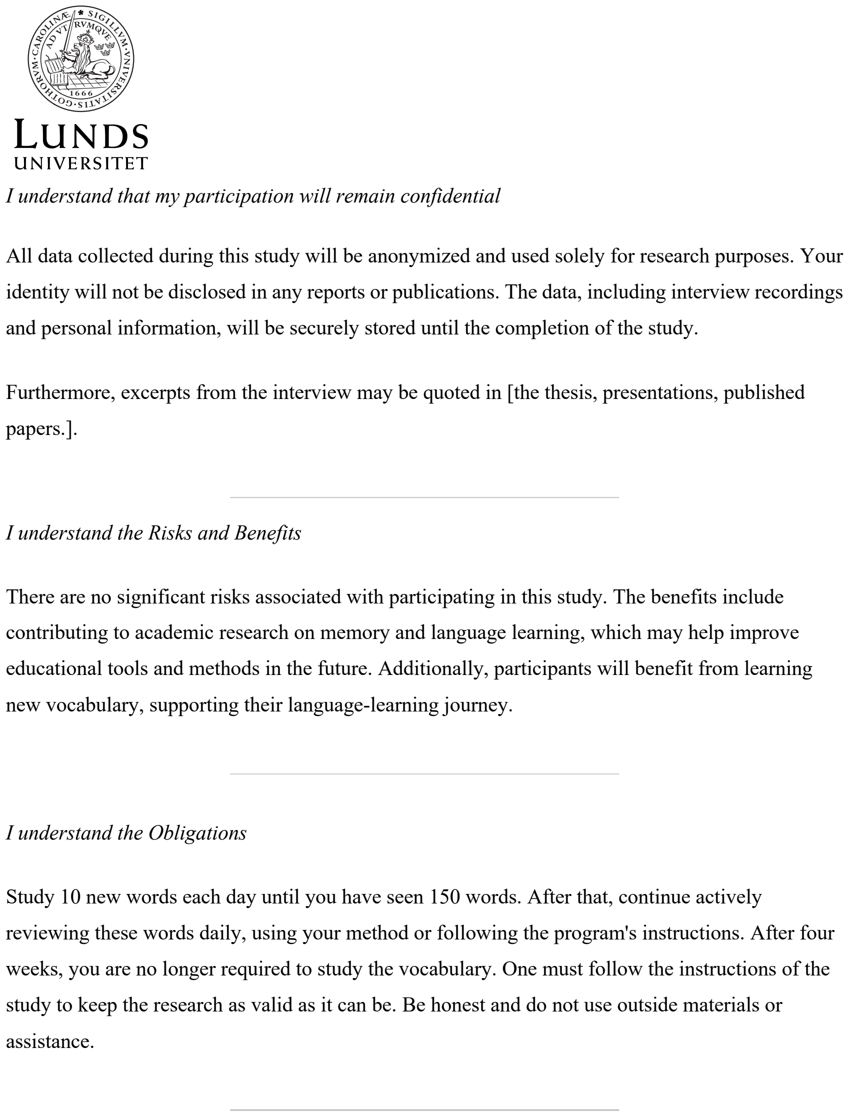
If you agree to participate, you will be asked to:

1. Either use a program to study vocabulary or follow your study method.
2. Participants will take part in a pre-interview at the start of the study, followed by a vocabulary assessment test.
3. Participants will later complete a post-test, followed by an additional test later. No preparation is required for these tests.
4. An additional post-interview will be conducted to gather additional insights.
5. Participants using the program will be asked to share the data generated from their usage. This does not apply to those using their study method.

I understand that this is a Voluntary Participation

Your participation in this study is entirely voluntary. You are free to refuse to participate or withdraw from the study at any time without any negative consequences.

Consent form page 1



I understand that my participation will remain confidential

All data collected during this study will be anonymized and used solely for research purposes. Your identity will not be disclosed in any reports or publications. The data, including interview recordings and personal information, will be securely stored until the completion of the study.

Furthermore, excerpts from the interview may be quoted in [the thesis, presentations, published papers.].

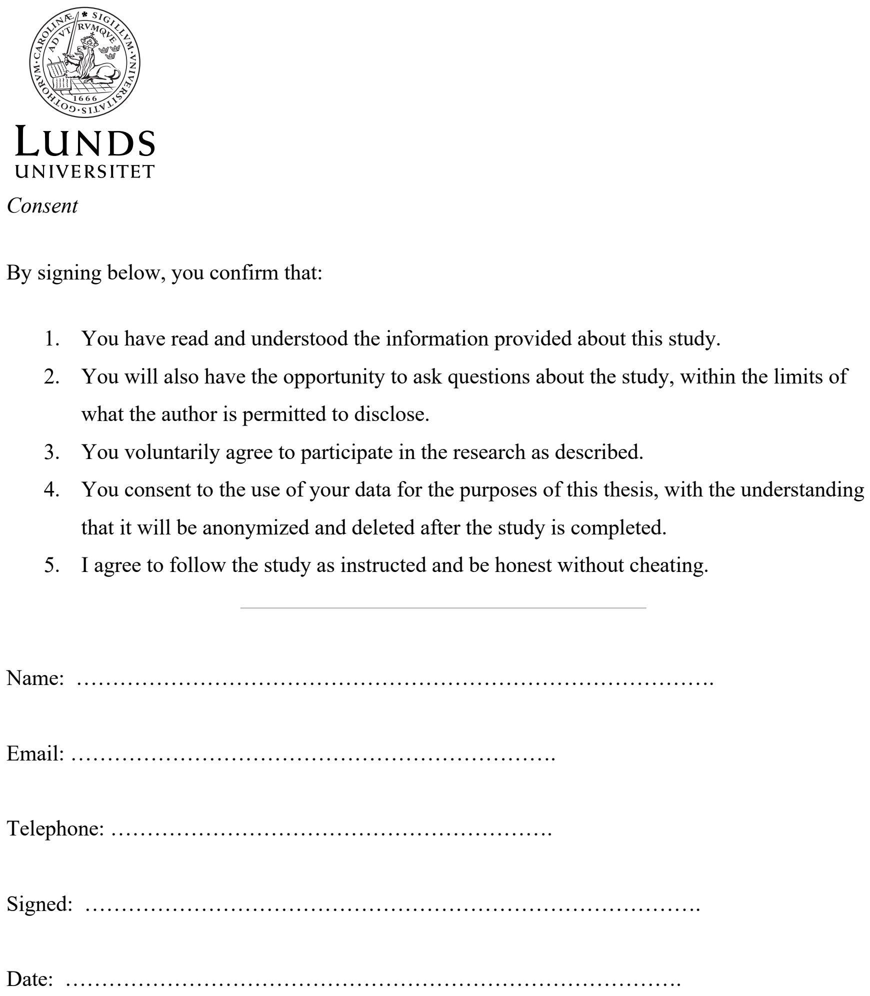
I understand the Risks and Benefits

There are no significant risks associated with participating in this study. The benefits include contributing to academic research on memory and language learning, which may help improve educational tools and methods in the future. Additionally, participants will benefit from learning new vocabulary, supporting their language-learning journey.

I understand the Obligations

Study 10 new words each day until you have seen 150 words. After that, continue actively reviewing these words daily, using your method or following the program's instructions. After four weeks, you are no longer required to study the vocabulary. One must follow the instructions of the study to keep the research as valid as it can be. Be honest and do not use outside materials or assistance.

Consent form page 2



LUNDS UNIVERSITET

Consent

By signing below, you confirm that:

1. You have read and understood the information provided about this study.
2. You will also have the opportunity to ask questions about the study, within the limits of what the author is permitted to disclose.
3. You voluntarily agree to participate in the research as described.
4. You consent to the use of your data for the purposes of this thesis, with the understanding that it will be anonymized and deleted after the study is completed.
5. I agree to follow the study as instructed and be honest without cheating.

Name: …………………………………………………………………………….

Email: ………………………………………………………….

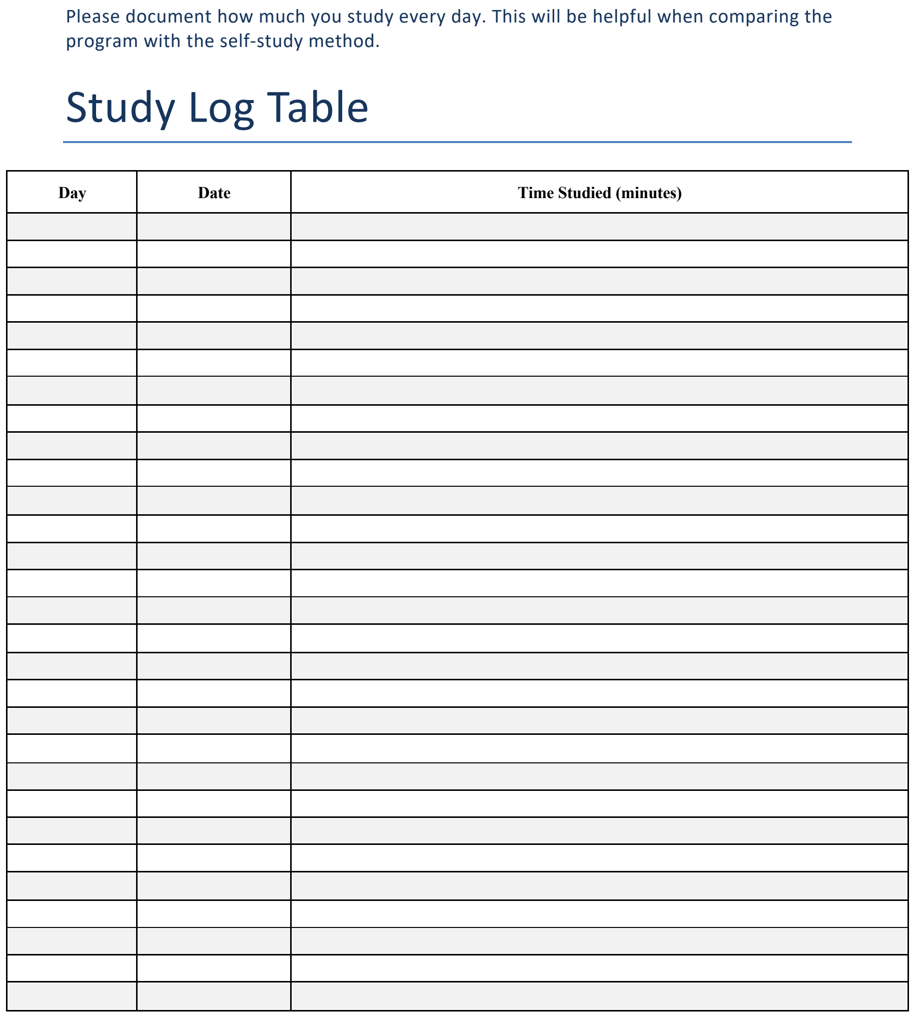
Telephone: …………………………………………………….

Signed: ………………………………………………………………………….

Date: ………………………………………………………………………….

Consent form page 3

Appendix B Study Log



Please document how much you study every day. This will be helpful when comparing the program with the self-study method.

Study Log Table

Day	Date	Time Studied (minutes)

Study Log

Appendix C Vocabulary List

だいがく 大学 university

りゅうがくせい 留学生 international student

せんこう 専攻 major

ともだち 友達 friend

にほん 日本 Japan

でんわ 電話 telephone

なまえ 名前 name

けいざい 経済 economics

こうがく 工学 engineering

せいじ 政治 politics

がっこう 学校 school

かいもの 買い物 shopping

いぬ 犬 dog

ねこ 猫 cat

ひと 人 person

こども 子供 child

しゃしん 写真 photograph

はな 花 flower

ごはん ご飯 meal

びょういん 病院 hospital

あね 姉 older sister

おとうと 弟 younger brother

きょうだい 兄弟 brothers and sisters

かいしゃ 会社 company

しょくどう 食堂 cafeteria

めがね 眼鏡 glasses

くるま 車 car

おなか お腹 stomach

ぶんがく 文学 literature

れきし 歴史 history

いしゃ 医者 doctor

しゅふ 主婦 housewife

べんごし 弁護士 lawyer

いもうと 妹 younger sister

かさ 傘 umbrella

さいふ 財布 wallet

しんぶん 新聞 newspaper

とけい 時計 watch

まえ 前 in front (of)

たべもの 食べ物 food

のみもの 飲み物 drink

くだもの 果物 fruit

りょこう 旅行 travel

うみ 海 sea

しゅくだい 宿題 homework

たんじょうび 誕生日 birthday

かんじ 漢字 Chinese character

おかね お金 money

ゆき 雪 snow

きおん 気温 temperature

ふゆ 冬 winter

かいしゃいん 会社員 office worker

しごと 仕事 job

こうりつ 効率 efficiency

とうせん 当選 winning an election

しじ 支持 support

ぼうし 帽子 hat

ゆうびんきょく 郵便局 post office

ちゅうごくじん 中国人 Chinese

えいご 英語 English

えいが 映画 movie

おんがく 音楽 music

ざっし 雑誌 magazine

おちゃ お茶 green tea

みず 水 water

いえ 家 house

にもつ 荷物 baggage

でんき 電気 electricity

でんしゃ 電車 train

くに 国 country

かぞく 家族 family

おじいさん お爺さん grandfather

おばあさん お婆さん old lady

ちち 父 father

はは 母 mother

あに 兄 older brother

でんとう 伝統 tradition

かつやく 活躍 taking an active part

れんらく 連絡 contact

ちほう 地方 Region

けいこう 傾向 tendency

ごうかく 合格 passing an examination

せんぞ 先祖 Ancestor

たいさく 対策 countermeasure

はれ	晴れ	sunny weather
あめ	雨	rain
しゅるい	種類	Type
でんりょく	電力	electric power
えいせい	衛生	hygiene
むりょう	無料	free of charge
しんりん	森林	forest
はかい	破壊	Destruction
さばく	砂漠	desert
ちょきん	貯金	money saved up
えんがん	沿岸	coast
いびき	鼾	snoring
げんりょう	原料	raw materials
はんとう	半島	peninsula
ちょくせつ	直接	directly
そうぞう	想像	imagination
かし	歌詞	song lyrics
きぼう	希望	hope
げんいん	原因	cause
でんせつ	伝説	folklore
にきび	面皰	Pimple
かいご	介護	Taking care of [Old People or Sick People]
せいさく	政策	policy
ろんぶん	論文	Thesis
わるぐち	悪口	bad-mouthing
かんし	監視	Monitoring
かせき	化石	fossil
けつぎ	決議	decision
あいまい	曖昧	ambiguous
ふこう	不幸	misfortune
しゅじんこう	主人公	protagonist
ほうりつ	法律	law
しゅっけつ	出血	bleeding
じょうしき	常識	common sense
へいき	兵器	weapon
しゅっさん	出産	childbirth
けいばつ	刑罰	(criminal) punishment
せきにん	責任	responsibility
れんたい	連帯	joint
りじゅん	利潤	profit
しょうめい	証明	proof
きゅうしゅう	吸収	absorption
いと	意図	intention
びじん	美人	beautiful woman
えがお	笑顔	Smile
しゅうかん	習慣	habit
きおく	記憶	memory
こてい	固定	fixed (in place)
しゅうしょく	就職	getting a full-time job
とうひょう	投票	vote
たいど	態度	attitude
とうろん	討論	debate
かんしゃ	感謝	gratitude
かっこ	括弧	parentheses
けが	怪我	injury
げんしょう	現象	phenomenon
かくとく	獲得	Acquisition
れんぽう	連峰	mountain range
ぎょうじ	行事	event
ないよう	内容	content
がか	画家	painter
きろく	記録	record
めいし	名刺	business card
げんきん	現金	cash
はんざい	犯罪	crime
じゆう	自由	freedom

All 150 Japanese English word pairs from the study