Analysis of Smartphone-Based Flashcard Apps for Second Language Vocabulary Acquisition

Taiga Kaitsu[1] and Tatsuya Nakata[2]

[1] Graduate School of Intercultural Communication, Rikkyo University, Japan

[2] College of Intercultural Communication, Rikkyo University, Japan

This manuscript was accepted for publication in *Computer Assisted Language Learning* (Informa UK Limited, Taylor & Francis Group) on March 14, 2025. This document may not exactly replicate the final version published in the journal.

The Version of Record of this manuscript has been published and is available in *Computer Assisted Language Learning* (March 24, 2025; https://www.tandfonline.com/doi/full/10.1080/09588221.2025.2481396)

Correspondence should be addressed to:

Tatsuya Nakata

College of Intercultural Communication, Rikkyo University

3-34-1 Nishi-Ikebukuro, Toshima-ku, Tokyo 171-8501 Japan

Email: nakata@rikkyo.ac.jp

Abstract

Flashcards are effective for vocabulary learning, which is crucial for second language (L2) acquisition. Digital flashcards may offer some specific advantages over paper flashcards. Although many flashcard apps for smartphones exist, some of them may not be completely designed to maximize L2 vocabulary learning. Accordingly, this study (a) developed a framework to analyze flashcard apps for smartphones based on 24 criteria elicited from current research on vocabulary acquisition, and (b) evaluated six commercially available flashcard apps (*Brainscape*, *iKnow!*, *Memrise*, *mikan*, *Quizlet*, and *WordHolic*) using the framework. The analysis showed that all six apps had multiple design features that facilitated vocabulary learning. Specifically, all six supported custom flashcard creation, offered multilingual support, included both presentation and retrieval modes, supported both receptive and productive recall, and gave feedback. At the same time, the results indicated that there may be further potential for improvement regarding flashcard creation and editing. Furthermore, the analysis affirmed that each app has its own strengths and weaknesses. For example, while *mikan* proved suitable for those who want to acquire various types of information about lexical items (e.g., inflections, derivations, associations), *Memrise* may be more beneficial for those who want to focus on spoken forms of vocabulary. These findings suggest that learners can and should choose apps based on their needs or preferences. This study will be valuable because it may provide useful guidelines for selecting, developing, and researching digital flashcard software for mobile technology.

Keywords: flashcards, vocabulary, smartphone apps, evaluation criteria, retrieval.

Introduction

Vocabulary knowledge constitutes an important aspect of second language (L2) acquisition (Nation, 2022). Learning from flashcards is considered one of the most effective strategies for acquiring vocabulary, which is important for language mastery (Webb et al., 2020). Flashcards are memorization tools that include an L2 lexical item on one side and its meaning (e.g., first language [L1] translation or L2 definition) on the other. Nation (2022) argues that flashcards are especially effective for learning the connections between form and meaning, among many aspects of vocabulary knowledge. Webb et al. (2020) conducted a meta-analysis that demonstrated the benefits of flashcard learning. Their analysis showed that flashcards led to larger vocabulary gains than word lists, writing words in sentences and compositions, or fill-in-the-blanks.

Nevertheless, flashcard learning is not without limitations. One common criticism is that although flashcards may be useful for learning the spelling and translation equivalent of L2 words, they do not necessarily facilitate knowledge about how to use the vocabulary. These impediments can be mitigated through meaning-focused activities, such as pleasure reading or extensive viewing of movies or TV shows. Empirical studies, furthermore, suggest that flashcard learning may result in knowledge that forms a basis for fluent, successful language use (Elgort, 2011; Elgort & Piasecki, 2014).

Today, digital and paper flashcards are widely used, although digital flashcards offer several advantages over paper flashcards (Ashcroft et al., 2018; Nakata, 2020; Teymouri, 2024; Xodabande et al., 2024). First, digital flashcards enable vocabulary learning through multimedia, such as images, videos, or audio. Second, they can provide exercises that are difficult to conduct using paper flashcards, such as multiple-choice questions or exercises with a time limit. Third, using mobile devices such as smartphones for L2 vocabulary learning may have positive effects on learners' motivation (Nikoopour & Kazemi, 2014; Xueting Ye & Shi, 2023; Zung et al., 2022). Fourth, digital flashcards are more portable than

paper flashcards, especially when learners want to learn a large number of words. Fifth, they may help implement effective learning strategies: many learners are unaware of effective strategies, such as distributed practice (Nakata & Suzuki, 2019; Zung et al., 2022). Digital flashcards can be programmed to help users learn vocabulary effectively, even if they are not aware of the strategies being employed.

Although many flashcard applications (apps) for smartphones exist, some of them may not have been fully developed to maximize vocabulary learning (Dunlosky & O'Brien, 2022; Lin et al., 2023). The purpose of this study, therefore, is to develop a framework for analyzing flashcard apps based on existing research and use this framework to evaluate commercially available apps. The findings of this study will be useful for language learners, teachers, researchers, and materials developers. Specifically, learners and teachers should find it easier to choose effective flashcard apps, while researchers will be equipped with a framework to investigate the flashcard apps. This will benefit the literature on computer-based flashcard learning. Finally, by proposing guidelines for designing effective flashcard apps, materials developers will acquire useful knowledge to create new flashcard apps or improve the existing ones.

The flashcard apps to be analyzed in this study were chosen based on criteria such as the average and number of user ratings on the app platforms in Japan (for details, see the Methods section), where flashcard learning is a popular vocabulary learning method (Schmitt, 1997). However, considering the tool's popularity in other countries such as the United States (Zung et al., 2022), the findings of this study may also be beneficial to learners based outside of Japan.

Literature Review

The study of computer-assisted flashcard programs is an evolving area of inquiry, and numerous studies have been published on this topic in the past few years (e.g., Barclay, 2024; Boroughani et al., 2023; Cygnet & Sivakumar, 2024; Dunlosky & O'Brien, 2022; Hsu

& Lin, 2024; Koleini et al., 2024; Lafleur, 2020; Lei & Reynolds, 2022; Lin & Lin, 2019; Muqaibal et al., 2023; Nakata et al., 2023; Serfaty & Serrano, 2024; Simonnet et al., 2025; van den Broek et al., 2023; Zung et al., 2022). However, most of these studies examined only one flashcard app, such as *Anki* (Koleini et al., 2024) or *Quizlet* (Boroughani et al., 2023; Hsu & Lin, 2024; Muqaibal et al., 2023; Serfaty & Serrano, 2024), without conducting an extensive assessment of the flashcard apps available for smartphones. A notable exception is the study by Dunlosky and O'Brien (2022), which analyzed 13 flashcard programs (e.g., *Anki*, *Brainscape*, and *Quizlet*), employing a framework derived from cognitive psychology. Their framework addressed five key aspects: test format, machine scoring, self-scoring, relearning, and materials. The analysis highlighted the strengths of the apps reviewed, including their capacity to enable learners to review previously studied items, incorporate pictures into flashcards, and share flashcards with other users. Although the findings of their study are valuable, their objective was not to analyze flashcard apps for L2 vocabulary learning but for general learning across subjects, such as psychology. For this reason, the evaluation criteria used in their study may not fully align with a framework to evaluate flashcard apps specifically for L2 vocabulary learning. Moreover, the apps analyzed by Dunlosky and O'Brien (2022) include both web- and smartphone-based apps, whereas this study focuses exclusively on smartphone-based apps.

Lin et al. (2023) also investigated 30 smartphone vocabulary learning apps using a framework derived from research on vocabulary learning and educational technology. Their framework addressed the following six aspects: learning tasks, goal clarity, feedback, strategy instruction, repetition, and instructional control. According to their findings, the top apps' characteristics include the capacity to give clear goals, and provide feedback to boost learning and motivation. Moreover, their study suggested that most programs could be improved by facilitating higher-order cognitive processes (i.e., applying, analyzing, evaluating, and creating), offering more scaffolded feedback, and providing various strategies, such as

contextual cues, semantic associations, word parts, and mnemonics, to facilitate the acquisition of the breadth and the depth of word knowledge. Although the findings of their study are useful, their primary objective was not to analyze flashcard apps but rather apps for vocabulary learning in general. Thus, the evaluation criteria used in their study do not necessarily help us develop a framework to evaluate flashcard apps, which is the focus of this study.

Nakata (2011) conducted a comprehensive investigation of flashcard programs for personal computers (PCs). Drawing on previous research, he proposed a framework consisting of 17 criteria to evaluate flashcard software. After analyzing nine commercially available flashcard programs, including *iKnow!*, *Quizlet*, and *SuperMemo*, using his framework, his findings suggested that most programs had several design features conducive to learning. Seven of the nine programs supported custom flashcard creation, offered multilingual support, and allowed learners to add various types of information (e.g., context, audio, or images) to flashcards. Nonetheless, Nakata also found that the programs had potential for improvement. Specifically, most were not designed to gradually increase exercise difficulty or provide opportunities to practice vocabulary in diverse contexts. These findings have provided useful guidelines for selecting, developing, and evaluating flashcard software.

However, there are at least three drawbacks to Nakata's (2011) study as well. First, his study analyzed flashcard programs for PCs. Considering that smartphones are now ubiquitous, it may be useful to conduct a similar study for flashcard apps for smartphones. The second limitation concerns the framework Nakata used for evaluating the flashcard software. Although his framework drew well on research that was available at the time of the study, given the recent developments in technology and L2 vocabulary acquisition research, it may be valuable to investigate additional features, such as automatic speech recognition (Bashori et al., 2024) or motivating feedback (Lin et al., 2023). Accordingly, the present

study attempted to improve the Nakata (2011) framework based on these developments. Consequently, the framework used in this study comprises 24 criteria (see the Methods section), a seven-item increase from Nakata's approach.

Another limitation of Nakata's (2011) study is that his criteria for selecting flashcard software for analysis were somewhat arbitrary. Of the nine programs he analyzed, only *P-Study System* and *Quizlet* were chosen based on objective criteria, i.e., the number of downloads and users. Therefore, the apps analyzed in this study were selected based on the clear quantitative criteria of the average and number of user ratings on app platforms (i.e., the App Store for iOS and Google Play for Android; see Methods for details).

Design Features for Effective Flashcard Apps

Proposing a framework to analyze flashcard apps necessitates identifying design features that facilitate flashcard learning. Accordingly, in this section, we review previous studies related to (1) flashcard creation and editing and (2) learning. The insights gained from this literature review will lay the groundwork for a framework to analyze the flashcard apps in the current study (for the evaluation framework, see the Methods section).

Flashcard Creation and Editing

Flashcard Creation. Many flashcard apps come with ready-made flashcards, such as those with lexical items related to travel, business, or medicine. Although ready-made flashcards may facilitate learning more than self-made flashcards (Dodigovic, 2013; Lei & Reynolds, 2022), they are not necessarily suitable for every learner. Ideal flashcard apps, therefore, should support custom card creation (Nakata, 2011).

Multilingual Support. Ideal flashcard apps should allow learners to use non-alphabet-based languages for both target items and their meanings (Nakata, 2011). Multilingual support for target items is useful because it allows users to study non-alphabet-based languages such as Chinese or Japanese. Multilingual support for the meanings would also allow learners whose L1 is non-alphabetical to use their L1 to establish initial form-

meaning mappings. Because the meanings of L1 translations may not necessarily correspond to those of L2 words, however, the use of L1 may be incomplete or misleading. Nevertheless, the use of L1 translation equivalents would at least be helpful for connecting the form and its primary meaning, since L1 use increases L2 vocabulary learning (e.g., Gyllstad et al., 2023). Multilingual support, therefore, is here considered a positive feature for flashcard apps.

Support for Multiple Words for Target Items and Their Meanings. Flashcard apps should allow learners to input multiple words for both target items and their meanings (Nakata, 2011). Support for multiple words for target items is useful because this allows learners to study multi-word units, such as idioms or phrasal verbs, which contribute to accuracy and fluency (Schmitt, 2023). The prominence of the multi-word units in language acquisition is also consistent with the usage-based approaches to language (Ellis & Wulff, 2020). Support for multiple words for meanings is also a positive feature because it allows learners to enter multiple meanings for polysemous items (e.g., "straight" and "right" for the Spanish *recto*).

Support for Flashcard Sets. Nakata (2011) argues that flashcard software should allow learners to create multiple flashcard sets. This is a useful feature because it allows learners to study items related to particular goals or topics (e.g., words related to travel, business, or medicine).

Sharing Flashcard Sets. The ability to share flashcard sets with other users would also be a beneficial feature (Dunlosky & O'Brien, 2022) because studying flashcards created by others will save learners from the task of making all their custom flashcards themselves. A sharing function also allows instructors to share flashcards with their students, ensuring that the students are exposed to vocabulary useful for their education.

Types of Information. Flashcard learning typically involves linking an L2 word form with its meaning, usually in the form of an L1 translation or an L2 definition. However, flashcard apps should also provide various kinds of information (such as audio, phonetic

symbols, images, videos, parts of speech, word parts, and example sentences or phrases) for at least two reasons. First, as pointed out by Nation (2022), learning L2 words involves much more than associating an L2 word form with its definition, and includes at least the following nine aspects: written form, spoken form, word parts, form-meaning connection, concepts and referents, associations, grammatical functions, collocations, and usage constraints. The ability to provide various types of information is useful because it facilitates the acquisition of multiple aspects of word knowledge. Second, the provision of visual information, such as images or videos, would also be useful because it may facilitate the retention of L2 vocabulary (Carpenter & Olson, 2012; Ramonda, 2022) according to the dual coding model (Paivio & Desrochers, 1980).

Learning

Presentation Mode. Research suggests that input constitutes the initial step for vocabulary learning (input-based incremental vocabulary instruction model; Barcroft, 2012). The flashcard apps, therefore, should support a presentation mode, in which learners are presented with information about the new target items, typically their forms and meanings.

Retrieval. Retrieval refers to the cognitive process in which learners are asked to recall or recognize information about target items. Retrieval is found to facilitate retention more than presentation, a phenomenon known as the retrieval practice (or testing) effect (Carpenter, 2023; Pan et al., 2024). Retrieval can be categorized based on two dichotomies: 1) receptive vs. productive, and 2) recognition vs. recall (Nakata, 2020). Receptive retrieval requires learners to retrieve meanings corresponding to L2 words, whereas productive retrieval requires learners to retrieve L2 words corresponding to meanings. Recall requires learners to generate a response (e.g., translate L2 words into L1, or vice versa), whereas recognition requires learners to respond without generating a response (e.g., choosing the correct answer from several options). The two dichotomies result in the following four

retrieval formats: receptive recognition, receptive recall, productive recognition, and productive recall (Nakata, 2020).

Studies suggest that productive and receptive retrieval are beneficial in different ways. One advantage of productive retrieval is its effectiveness; productive retrieval often yields greater productive knowledge than receptive retrieval, while at the same time leading to comparable gains in receptive knowledge (e.g., Griffin & Harley, 1996; Nakata, 2016; Schneider et al., 2002; Webb, 2009). One advantage of receptive retrieval, in contrast, is that it requires less time and effort (Nation & Webb, 2011). Furthermore, receptive retrieval has been found to be more effective than productive retrieval for lower-proficiency learners (Terai et al., 2021). As both productive and receptive retrieval offer benefits, flashcard apps should support both receptive and productive retrieval. Regarding the recognition-recall dichotomy, recall often leads to better retention than recognition (Technique Feature Analysis; Nation & Webb, 2011), whereas recognition is more efficient than recall (Nakata, 2016). Thus, flashcard apps should also support both recognition and recall.

Varied Encounters and Use. Varied encounters and use refer to encountering or using partially known lexical items in new or different contexts than before (e.g., different meanings, collocations, inflections, etc.), and help deepen learners' understanding of lexical items (Webb & Nation, 2017). For instance, suppose that a learner first encounters the verb "make" in the sentence "Jack makes toys." They might hypothesize that "make" is only used for creating physical objects. Exposure to sentences such as "I *made* a reservation," "They will *make* a great couple," or "Don't *make* a promise if you cannot keep it." (i.e., varied encounters) would thus help deepen their understanding of the verb "make." The degree of variation also affects vocabulary learning (Technique Feature Analysis; Nation & Webb, 2011). The degree of variation is high when the same lexical item is used in different contexts with different meanings, collocations, or inflections compared with previous encounters. Conversely, the degree is low when it is used with similar meanings, collocations, or

inflections. For instance, let us assume a learner first encounters the English word "round" in the phrase "a *round* face." The degree of variation is low when the second encounter involves the phrase "*round* eyes" because the meaning is similar (i.e., "having a shape like a circle"), and the part of speech in both is the same (i.e., adjective). In contrast, when the second encounter involves the phrase "a *round* of drinks," the degree of variation is high because it is used with a different meaning (i.e., "a series or cycle"), and the part of speech is different (i.e., noun).

Increasing Exercise Difficulty. The desirable difficulty framework (Suzuki et al., 2020), according to which effortful but successful retrieval enhances retention, suggests that flashcard software should be designed to gradually increase exercise difficulty for a given target item. Exercise difficulty can be gradually increased by offering receptive retrieval before productive retrieval, or recognition before recall.

Block Size. Block size is defined as the number of items that learners study at a time (Nakata & Webb, 2016). For instance, if a learner studies 10 items at a time in a flashcard set consisting of 100 items, the block size is 10. Existing studies have yielded inconsistent results regarding block size. While several have found an advantage of larger over smaller block sizes (e.g., Nakata & Suzuki, 2019; Nakata et al., 2023), Nakata and Webb (2016) found that block size may have little effect on retention when temporal spacing is controlled (distributed practice effect; Kim & Webb, 2022). Because of these mixed results, flashcard apps should ideally offer flexible block sizes to allow for user adjustment based on individual needs and preferences.

Interference Avoidance. Semantic clustering refers to learning semantically related words (e.g., names of animals, colors, or food) simultaneously (Nation & Webb, 2011). Despite the widespread use of semantic clustering, vocabulary researchers have warned against this practice because semantic clustering may lead to interference errors (Nation & Webb, 2011). For instance, learners who are presented with the Japanese words *inu* ("dog")

and *neko* (“cat”) at once could experience cross-association, which would result in the learners mistakenly associating *inu* with “cat” and *neko* with “dog” (interference theory). Whereas earlier research has demonstrated the negative effects of semantic clustering (Papathanasiou, 2009; Wilcox & Medina, 2013), recent research suggests that the effects of semantic clustering may not be very robust after all (e.g., Kemp & McDonald, 2021; Nakata & Suzuki, 2019). Nevertheless, although semantic clustering may not necessarily inhibit long-term retention, it may still have negative effects in the short term (Bach & Barclay, 2025; Nakata & Suzuki, 2019). The ability to avoid interference is therefore a positive feature. This can be achieved by not including semantically related items in a given flashcard set.

Fluency Development. Fluency refers to the speed at which learners understand or produce language. Facilitating fluency would be a positive feature among flashcard apps because fluent access to lexical knowledge is essential for successful communication (four strands; Nation, 2022). Fluency development can be incorporated into flashcard apps by imposing a time limit on retrieval practice.

Automatic Speech Recognition. Bashori et al. (2024) found that automatic speech recognition, in which the app provides feedback in response to the learner’s pronunciation, facilitates the acquisition of the spoken forms of L2 words. The ideal flashcard apps, therefore, should support automatic speech recognition.

Adaptive Sequencing. Adaptive sequencing refers to an algorithm that automatically adjusts the practice schedule for target items based on learner performance (Nakata, 2020). This would be a positive feature because it allows learners to focus on difficult or unknown items instead of easy or already known items.

Retirement Criterion. The retirement criterion is concerned with when target items are eliminated from further practice and can be categorized into the following three types: a dropout schedule, a fixed schedule, and self-judgment. In a dropout schedule, target items

that are answered correctly a certain number of times (e.g., once or twice) are automatically dropped from further study (Nakata et al., 2023). In a fixed schedule, target items are practiced for a fixed number of times (e.g., four times) regardless of the number of correct responses (Yamagata et al., 2023). In self-judgment, target items are dropped from further practice based on learners' evaluation of their own understanding or confidence (Ariel & Karpicke, 2018). Existing studies have yielded conflicting results regarding the effects of retirement criteria. While some studies have shown an advantage of dropout schedules (Pyc & Rawson, 2007), others have failed to do so (Kornell & Bjork, 2008; Pyc & Rawson, 2011). Although the optimal retirement criterion is unclear and perhaps depends on several item-related or learner-related factors (e.g., the difficulty of target items or the working memory capacity of learners), it would be useful to investigate what kinds of retirement criteria are common among flashcard apps.

Expanded Rehearsal. Expanded rehearsal (also known as *expanding spacing*) refers to a practice schedule in which time intervals for a given target item are gradually increased (Leonard et al., 2024). Although some researchers argue that expanded rehearsal enhances vocabulary learning (expanded retrieval effect; Nation & Webb, 2011), recent research suggests that it may not necessarily do so (e.g., Kim & Webb, 2022). Nonetheless, Nakata (2011) has found that expanded rehearsal is a common feature among flashcard software; seven out of nine flashcard programs analyzed by him supported expanded rehearsal. Accordingly, expanded rehearsal will be investigated to ascertain whether it is a common feature among flashcard apps for smartphones.

Feedback. Feedback refers to information such as the correct answer or the knowledge of results (e.g., correct or incorrect) given in response to learner performance (Lin et al., 2023). The provision of feedback is a positive feature because it facilitates learning (Nakata, 2020). Flashcard apps should also provide feedback that is designed to motivate learners (Lin et al., 2023). Examples of such feedback include comments such as "Great job!"

at the end of a learning session. Another example would be the provision of user rankings, which may encourage learners to study harder so they will be highly ranked. This kind of feedback is a desirable feature because it can have positive effects on learner motivation (Mishra, 2006).

Present Study

The aim of this study is twofold; first, to develop a framework to analyze flashcard apps for smartphones based on the reviewed literature, and second, to use this framework to evaluate commercially available flashcard apps while addressing the limitations of the past research (Dunlosky & O'Brien, 2022; Lin et al., 2023; Nakata, 2011). The research question for this study is as follows: What are the common strengths and weaknesses of existing smartphone-based flashcard apps for L2 vocabulary learning?

Methods

Inclusive Criteria for Flashcard Apps

The flashcard apps under investigation were selected based on the following four criteria:

1. The app is compatible with both iOS and Android.
2. The app has an average user rating of 4.0 or higher out of 5.0 on both the App Store (iOS) and Google Play (Android) in Japan.
3. The app has received a combined total of 10,000 or more user ratings on the App Store and Google Play.
4. The app was multipurpose and not limited to specific purposes, such as preparation for TOEFL or IELTS.

Based on the above criteria, the following six apps were chosen: *Brainscape*, *iKnow!*, *Memrise*, *mikan*, *Quizlet*, and *WordHolic*. Details of the apps are presented in Table 1.

<Insert Table 1 here>

Evaluative Criteria for the Flashcard Apps

Drawing on previous research reviewed in the Literature Review, a framework for analyzing flashcard apps, which consists of 24 criteria, was developed. The first seven criteria were concerned with flashcard creation and editing, and the other 17 pertained to learning:

Flashcard Creation and Editing

1. Does it allow users to create custom flashcards?
2. Does it offer multilingual support?
3. Does it allow users to input multiple words for target items and their meanings?
4. Does it allow users to create multiple sets of flashcards?
5. Does it allow users to share flashcard sets with other users?
6. What kinds of lexical information does it provide for ready-made flashcards?
7. What kinds of lexical information does it allow learners to add to custom flashcards?

Learning

8. Does it offer opportunities for presentation?
9. Does it offer opportunities for retrieval?
10. Does it offer opportunities for receptive recognition?
11. Does it offer opportunities for receptive recall?
12. Does it offer opportunities for productive recognition?
13. Does it offer opportunities for productive recall?
14. Does it offer opportunities for varied encounters and use?
15. Does it gradually increase exercise difficulty?
16. Does it allow users to choose the block size?
17. Does it allow users to avoid interference from semantic clustering?
18. Does it offer opportunities for fluency development?
19. Does it offer automatic speech recognition?
20. Does it offer adaptive sequencing?
21. What is the criterion for dropping target items from further practice?

22. Does it offer expanded rehearsal?
23. Does it offer feedback for learning?
24. Does it offer motivating feedback?

Procedure

The first author of this study analyzed the six apps using the 24 criteria as described above. When analyzing complex features like algorithms, the official websites for each app were also consulted. For Criterion 6 (Types of information for ready-made flashcards) and Criterion 7 (Types of information for custom flashcards), we examined whether the following types of information were supported: L2 definitions, L1 translations, audio, phonetic symbols, images, videos, parts of speech, word parts, inflections, derivations, associations, example sentences or phrases, and user notes. Note that for Criterion 7, we only analyzed the ready-made flashcards created by the app developers and did not include those created by other users. This is because the types of information provided vary depending on the creators of flashcards.

For each criterion, + was awarded if the app met the criterion, whereas − was awarded if it did not, except for the following: Criteria 6, 7, 14, 15, 16, and 21. For Criterion 6 (Types of information for ready-made flashcards), the coding scheme included N/A in addition to + and –. The abbreviation N/A was assigned if the developers of the apps did not provide the ready-made flashcards. For Criteria 7, 14, 15, and 16, the coding scheme included ++ in addition to + and –. For Criterion 7 (Types of information for custom flashcards), ++ was assigned if the app supported data entry by automatically supplying it (e.g., when the learners entered a target word, the app automatically supplied its definition or parts of speech). The symbol + was assigned if learners needed to add the information themselves, whereas − was assigned if the app did not support the particular type of information.

For Criterion 14 (Varied encounters and use), ++ was awarded if the app was designed to facilitate a high degree of variation, whereas + was awarded if the app facilitated only a

low degree of variation. The degree of variation was rated as high when the same lexical item was used with different meanings, collocations, or inflections compared to the previous encounters, whereas it was deemed low when it was used in similar contexts with similar meanings, collocations, or inflections. For Criterion 15 (Increasing exercise difficulty), ++ was awarded if the app automatically arranged the exercises in order of difficulty for a given target item. The symbol + was awarded if the learners could increase the difficulty by choosing the exercises order themselves. The symbol − was awarded if the app was not designed to increase the exercise difficulty. For Criterion 16 (Block size), ++ was given if any block size was available. The symbol + was assigned if the learners could select the block size from several options, whereas − was given if the learners could not select the block size. For Criteria 21 (Retirement criterion), the retirement criterion was coded as *self-judgment* if the target items were dropped from further practice based on learners' self-evaluation of understanding or confidence. The criterion was coded as *auto-drop 1* if the target items were automatically dropped from further practice after one correct response. The coding *auto-drop* was used if the target items were automatically dropped from further practice after a certain number of correct responses, but the number of correct responses required was not disclosed to users.

To enhance the reliability of the analysis, developers from the six apps were invited to verify the results. The developers of *iKnow!*, *Memrise*, and *mikan* accepted the invitation and confirmed the accuracy of the analysis for their respective apps. To identify which app had the largest number of design features to facilitate vocabulary learning, a total score was calculated after the coding was completed. For Criteria 7, 14, 15, and 16, ++ was awarded one point, + was awarded 0.5 points, and − was awarded 0 points. Criteria 21 (Retirement criterion) and 22 (Expanded rehearsal) were excluded from the score calculation because existing studies have been inconclusive regarding the optimal retirement criterion or the effects of expanded rehearsal (see the Literature Review). For the other criteria, + was

awarded one point, and both − and N/A were awarded 0 points. The maximum possible score was 46.

Results

The results of the analysis are presented in Table 2. The table shows that the six apps have 13 strengths and three weaknesses in common. The common strengths were flashcard creation, multilingual support, support for multiple words for target items and their meanings, support for flashcard sets, L2 definitions and L1 translations in custom flashcards, presentation mode, retrieval mode, receptive recall, productive recall, increasing difficulty, interference avoidance, and feedback for learning. In contrast, the common weaknesses were the lack of the following features: L2 definitions for ready-made flashcards, user notes for ready-made flashcards, and videos for custom flashcards. Table 2 also shows that the six apps were ranked in descending order according to their total scores as follows: *mikan*, *iKnow!*, *Memrise*, *Quizlet*, *Brainscape*, and *WordHolic.*

<Insert Table 2 here>

The following text provides detailed information on the criteria that warrant further explanation. Regarding Criterion 6 (Types of information for ready-made flashcards), Table 2 shows that images were provided only by *iKnow!*, and videos were provided only by *Memrise*. In *iKnow!*, the images related to the meanings of the target words accompanied the ready-made flashcards (Figure 1). *Memrise* provided multiple videos for each vocabulary item, where different L1 speakers pronounce the same item with different accents or speeds in each video (Figure 2).

<Insert Figures 1 and 2 here>

As for Criterion 7 (Types of information for custom flashcards), *WordHolic* was the only app that supported the user notes for custom flashcards, where the learners could add any kind of text (e.g., words, phrases, or sentences) to the flashcards they created. Figure 3 shows a custom flashcard for the English word “produce” with various types of information,

such as frequency information from corpora, its word family members (i.e., inflections and derived forms), verb meaning, noun meaning, and example sentences.

<Insert Figure 3 here>

Table 2 also shows that only *iKnow* and *mikan* supported varied encounters and use (Criterion 14). In *iKnow*, this was achieved by providing multiple sentences for a given target word. For instance, the following two sentences were provided for the English word "produce" in *iKnow*:

We *produce* all our products here.

Our plan *produced* excellent results.

Encountering and using the same word in multiple contexts may result in varied encounters and use. However, it should be noted that *iKnow* is designed to facilitate only a low degree of variation. This is because each flashcard addresses only one part of speech and meaning for a given word. For instance, as shown previously, the flashcard for "produce" gives the verb meaning, but not the noun meaning (e.g., "fresh produce" or "agricultural produce"). Furthermore, although the verb "produce" is polysemous, only one meaning is addressed (i.e., "to create or make"). Other meanings of the verb, such as "to present or show" (e.g., "produce one's driver's license"), are either not addressed or included on separate flashcards. *mikan* facilitates varied encounters and use by creating a link from a target word to a dictionary entry. As shown in Figure 4, clicking the target word "produce" opens a dictionary entry. The entry includes both the verb (e.g., "produces electric cars") and noun meanings (e.g., "dairy produce"), as well as multiple meanings and example sentences for each part of speech. The findings suggest that although both *iKnow* and *mikan* facilitate varied encounters and use, *mikan* may lead to a higher degree of variation.

<Insert Figure 4 here>

The analysis also shows that *iKnow* and *Memrise* were the only apps designed to automatically increase the difficulty of the exercises as learning progressed (Criterion 15). This is achieved by presenting the exercises in the order of their difficulty. *iKnow* does this by first testing a given target word in a receptive recognition format (Figure 5, top), followed by productive recognition (Figure 5, middle) and productive recall formats (Figure 5, bottom), thus gradually increasing difficulty.

<Insert Figure 5 here>

Regarding Criterion 17 (Interference avoidance), all apps were awarded +. Interference resulting from semantic clustering can be avoided in two ways. First, custom flashcard sets can be created in which semantically related lexical items (e.g., names of animals, colors, or food) are not introduced simultaneously. Alternatively, ready-made flashcard sets that do not involve semantic clustering can be used. Notably, although most ready-made flashcard sets prepared by the app developers do not involve semantic clustering, a small number of them introduce semantically related lexical items together. Figure 6, for instance, shows a ready-made flashcard set offered by *iKnow!*, where names of animals (e.g., "cheetah," "elephant," and "tiger") are introduced together, demonstrating semantic clustering.

<Insert Figure 6 here>

Table 2 also indicates that *Memrise* was the only app that offered automatic speech recognition (Criterion 19). With this feature, users are asked to pronounce a target lexical item ("mover" in Figure 7), and the app provides feedback in response to the learner's pronunciation. However, the automatic speech recognition feature in *Memrise* is limited in the sense that the feedback indicates only whether a target word is pronounced correctly ("Sounds good!" for a correctly pronounced word and "Mmm...not quite!" for other responses). Regrettably, more detailed feedback for each syllable or phoneme is unavailable.

<Insert Figure 7 here>

Discussion

This study investigated the strengths and weaknesses of flashcard apps on smartphones. The results showed that all six apps share 13 strengths, which suggests that most flashcard apps for smartphones have been developed to facilitate vocabulary learning. Five out of the 13 strengths (i.e., flashcard sets, presentation mode, retrieval mode, receptive recall, and productive recall) have also been reported by Nakata (2011) as common strengths of flashcard programs for PCs. This suggests that the strength of flashcard programs for both PCs and smartphones lies in their ability to support flashcard sets, offer both presentation and retrieval modes, and support both receptive and productive recall. Among these five features, existing studies suggest that the presentation mode, retrieval mode, receptive recall, and productive recall are particularly beneficial for L2 vocabulary learning.

First, all six flashcard apps examined in this study offered the presentation mode, in which learners are presented with information about new target items (e.g., written forms, L1 translations, L2 definitions). This is a useful feature according to the input-based incremental vocabulary instruction model (Barcroft, 2012), which states that input constitutes the initial step for vocabulary learning. Second, in all six flashcard apps examined, the presentation mode was followed by the retrieval mode, which requires learners to recall or recognize information about target items. This is another useful feature because retrieval facilitates retention more than presentation, a phenomenon known as the retrieval practice (or testing) effect (Carpenter, 2023; Pan et al., 2024). Third, this study showed that all six flashcard apps supported receptive and productive recall. The support for both directions of retrieval is useful because studies suggest that receptive and productive retrieval are beneficial in different ways (see the Literature Review).

The results of this study also showed that the six apps had three weaknesses in common. Specifically, none supported (a) L2 definitions for ready-made flashcards, (b) user notes for ready-made flashcards, or (c) videos for custom flashcards. Significantly, all these

identified weaknesses are concerned with flashcard creation and editing (Criteria 1 to 7), rather than learning (Criteria 8 to 24). The findings suggest that although most flashcard apps have been developed to facilitate learning, there may be room for improvement regarding flashcard creation and editing. The lack of the L2 definitions in ready-made flashcards may not be a major problem considering that studies suggest that L1 translations lead to larger vocabulary gains than L2 definitions (e.g., Gyllstad et al., 2023). However, it may be useful to address the lack of the other two features, namely, the user notes in the ready-made flashcards and videos in custom flashcards. First, the lack of support for user notes in the ready-made flashcards is unfortunate as user notes would allow learners to add any kind of text, such as keywords or loanwords, which may help vocabulary learning (Nakata, 2020; Nation 2022). Second, none of the six flashcard apps examined in this study allowed the learners to add videos to the custom flashcards. Video support would be useful because visual information can facilitate the retention of L2 vocabulary according to the dual coding model.

Strengths and Weaknesses of Each Flashcard App

The six apps were ranked in descending order according to their total scores as follows: *mikan*, *iKnow!*, *Memrise*, *Quizlet*, *Brainscape*, and *WordHolic. mikan*'s strength is its ability to support various types of lexical information, such as phonetic symbols, parts of speech, inflections, derivations, associations, and example sentences or phrases in both ready-made and custom flashcards. Furthermore, *mikan* was also the only app to support information about inflections, derivations, and word parts for both ready-made and custom flashcards. The findings suggest that this app may be especially suitable for acquiring multiple aspects of word knowledge.

iKnow!, which Nakata (2011) regarded as the best program, received the second-highest score in this study (30.0), after *mikan* (35.0). The strengths of *iKnow!* lay in its ability to facilitate learning as it received the highest score (13.0) for learning-related criteria, along with *Memrise* and *mikan* (Table 2). Concomitantly, it should be noted that *mikan* and *iKnow!*

are targeted at the relatively narrow demographic of Japanese-speaking English learners. In contrast, the other four apps target a larger user base (see Table 1). The relatively high scores of *mikan* and *iKnow!* may be partially attributed to this narrow focus. In other words, because these apps were designed for a specific demographic, it may have been easier for developers to create ready-made flashcards containing various types of information (Criterion 6), or to support data entry for custom flashcards (Criterion 7), thereby resulting in higher scores. These findings suggest a potential tradeoff for developers between the app effectiveness and potential revenue streams. Targeting a specific demographic and maintaining a narrow focus may enable developers to create effective apps. However, from a financial standpoint, targeting a broader user base may be more desirable because it provides more opportunities for revenue generation.

Memrise received the third-highest total score (23). This app is unique in that it is the only one that offers videos for ready-made flashcards. Multiple videos are available for each vocabulary item, and different L1 speakers pronounce the same item in different accents or speeds in each video, which may facilitate the learning of spoken forms (Uchihara et al., 2022). In addition, *Memrise* is the only app that offers automatic speech recognition, which may also be beneficial for learning productive knowledge of spoken forms (Bashori et al., 2024). However, this feature in *Memrise* indicates only whether a target word is pronounced correctly (Figure 7). It would be more useful if the app could provide more detailed feedback for each syllable or phoneme in the target word.

Quizlet received the fourth-highest score (19.5). The app is unique in offering all four retrieval formats—receptive recognition, productive recognition, receptive recall, and productive recall—not only in its paid version but also in its free version. The four retrieval formats are also available in *mikan*, *iKnow!*, and *Memrise*. However, the free version of *mikan* offers only receptive recognition, and *iKnow!* and *Memrise* are not free apps (Table 1).

Therefore, *Quizlet* is beneficial for learners who want to practice all retrieval formats free of charge.

Brainscape received the fifth-highest total score (17). One significant feature of the app that is not covered in Table 2 is that it offers ready-made flashcards for a wide range of subjects, including world history, driver's licenses, math, and medical science. *Brainscape* may therefore be especially beneficial for learners who want to study other subjects alongside L2 vocabulary.

Although *WordHolic* received the lowest score (16.5), it is unique in that it is the only app that supports user notes for custom flashcards. This means that learners can add any kind of text (e.g., words, phrases, or sentences) to the flashcards they create (Figure 3). This feature may compensate for some of the app's weaknesses (e.g., lack of support for word parts, inflections, or derivations) because learners can add various types of information, including word parts, inflections, or derivations, as user notes. Learners can also add mnemonics such as keywords or loanwords as user notes, which may help vocabulary learning (Nakata, 2020; Nation 2022).

Conclusion

Although various flashcard apps for smartphones exist, not all of them may have been fully developed to maximize vocabulary learning, which underscores the need to establish a framework for analyzing flashcard apps, grounded in existing research. Such frameworks are beneficial as they can assist learners and teachers in selecting effective flashcard apps and offer researchers with a structure for investigating these apps. Furthermore, by proposing guidelines for designing effective flashcard apps, these frameworks can support materials developers in creating new flashcard apps or improving the existing ones. Thus, in this study, we developed a framework for analyzing flashcard apps for smartphones based on 24 criteria elicited from existing research. Six commercially available flashcard apps were evaluated using the framework. The findings showed that all six apps have 13 common strengths, which

suggests that most flashcard apps for smartphones have been developed to facilitate vocabulary learning. At the same time, the study also suggested potential areas for the improvement of the apps surveyed, particularly in the areas of flashcard creation and editing (Criteria 1 to 7).

Some flashcard apps scored higher than others, but each app had its strengths and weaknesses. For instance, *mikan* may be suitable for those who want to learn various types of information about lexical items (e.g., parts of speech, inflections, derivations, associations), whereas *Memrise* may be beneficial for those who want to learn spoken forms. Furthermore, *Quizlet* may be useful for those who want to practice four retrieval formats free of charge, whereas *WordHolic* may suit those who want to learn vocabulary using mnemonics (e.g., keywords or loanwords). These findings suggest that deciding on the most effective app may not be a straightforward process. Therefore, it is most beneficial for learners to choose an app that meets their needs or preferences.

The findings of this study are valuable because of the theoretical guidelines they provide for selecting, developing, and evaluating flashcard software. While this information is instructive, the present study also has several limitations. First, existing studies are still inconclusive regarding the optimal way to learn from flashcards. The evaluation framework used in this study therefore contains some inherent uncertainties. For instance, as pointed out in the Literature Review, it is still unclear whether semantic clustering (Criterion 17) and expanded rehearsal (Criterion 22) have positive or negative effects on vocabulary learning. Furthermore, existing studies have yielded inconsistent results regarding the optimal block size (Criterion 16) or retirement criterion (Criterion 21). Future research on L2 vocabulary development would allow researchers to further refine the evaluation criteria.

Second, given that smartphone apps are updated frequently, similar analyses may produce different results in the future. Teachers, learners, and researchers need to be cognizant of the latest features of these apps. Third, due to the relatively narrow focus of this

study, some popular apps had to be excluded from the analysis. For instance, *Duolingo*, which has more than 20 million user ratings on Google Play, was not analyzed because it is not a flashcard app. Similarly, *Eiken Eitango*, a popular app for learning English vocabulary for the Eiken Exam in Japan with more than 10 million user ratings on Google Play, was also excluded because it is designed specifically for the preparation for the exam and is not a multipurpose app (see Inclusive Criteria for Flashcard Apps in the Methods). Lastly, the apps analyzed in this study were chosen from the App Store and Google Play in Japan. In future research, it may be useful to select apps from platforms in other countries. Considering the value of flashcard learning and the benefits afforded by flashcard apps, further research investigating flashcard apps will be a useful follow-up to this study.

Acknowledgments

This work was supported by the JSPS KAKENHI under Grant 22K00743 and the Rikkyo University Special Fund for Research. The authors would like to thank Ron Martin, Satomi Mori, Ayako Aizawa, Sayane Soma, and Wai Yip Ho for their invaluable advice.
The manuscript has been proofread by Enago (www.enago.jp). The authors have used ChatGPT (GPT-3.5 and GPT-4) for language improvement. No generative AI was used for other purposes, such as idea generation, content generation, or analysis.

Disclosure Statement

The authors declare no conflicts of interest. They have neither affiliation with nor financial interest in any of the apps analyzed in this study.

References

Ariel, R., & Karpicke, J. D. (2018). Improving self-regulated learning with a retrieval practice intervention. *Journal of Experimental Psychology: Applied*, *24*(1), 43–56. https://doi.org/10.1037/xap0000133

Ashcroft, R. J., Cvitkovic, R., & Praver, M. (2018). Digital flashcard L2 vocabulary learning out-performs traditional flashcards at lower proficiency levels: A mixed-methods study of 139 Japanese university students. *The EuroCALL Review*, *26*(1), 14–28. https://doi.org/10.4995/eurocall.2018.7881

Bach, N. T. A., & Barclay, S. (2025). The effect of semantic and physical similarity on vocabulary learning. *The Language Learning Journal*, *53*(1), 40–52. https://doi.org/10.1080/09571736.2023.2286239

Barclay, S. (2024). The effect of form presentation mode and language learning aptitude on the learning burden and decay of L2 vocabulary knowledge. *International Review of Applied Linguistics in Language Teaching*. Advance online publication. https://doi.org/10.1515/iral-2023-0165

Barcroft, J. (2012). *Input-based incremental vocabulary instruction*. TESOL.

Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2024). 'Look, I can speak correctly': Learning vocabulary and pronunciation through websites equipped with automatic speech recognition technology. *Computer Assisted Language Learning, 37*(5–6), 1335–1363. https://doi.org/10.1080/09588221.2022.2080230

Boroughani, T., Behshad, N., & Xodabande, I. (2023). Mobile-assisted academic vocabulary learning with digital flashcards: Exploring the impacts on university students' self-regulatory capacity. *Frontiers in Psychology*, *14*, 1112429. https://doi.org/10.3389/fpsyg.2023.1112429

Carpenter, S. K. (2023). Encouraging students to use retrieval practice: A review of emerging research from five types of interventions. *Educational Psychology Review*, *35*(4), 96. https://doi.org/10.1007/s10648-023-09811-8

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

Cygnet, A. B., & Sivakumar, P. (2024). Efficacy of augmented reality-based flashcards on learning Basic Tamil words among primary learners during neo-normal period. *Education and Information Technologies, 29*(6), 6705–6722. https://doi.org/10.1007/s10639-023-12073-0

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

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning*, *61*(2), 367–413. https://doi.org/10.1111/j.1467-9922.2010.00613.x

Elgort, I., & Piasecki, A. E. (2014). The effect of a bilingual learning mode on the establishment of lexical semantic representations in the L2. *Bilingualism: Language and Cognition*, *17*(3), 572–588. https://doi.org/10.1017/S1366728913000588

Ellis, N. C., & Wulff, S. (2020). Usage-based approaches to L2 acquisition. In B. VanPatten, G. Keating, & S. Wulff (Eds.), *Theories in second language acquisition: An introduction* (pp. 63–82). Routledge.

Griffin, G. F., & Harley, T. A. (1996). List learning of second language vocabulary. *Applied Psycholinguistics*, *17*(4), 443–460. https://doi.org/10.1017/S0142716400008195

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

Hsu, H., & Lin, C. (2024). Factors influencing students' listening learning performance in mobile vocabulary-assisted listening learning: An extended technology acceptance model. *Journal of Computer Assisted Learning*, *40*(4), 1511–1525. https://doi.org/10.1111/jcal.12969

Kemp, L. S., & McDonald, J. L. (2021). Second language vocabulary acquisition: The effects of semantic relatedness, form similarity, and translation direction. *Language Learning*, *71*(3), 716–756. https://doi.org/10.1111/lang.12449

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

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*, 100344.

Kornell, N., & Bjork, R. A. (2008). Optimising self-regulated study: The benefits—and costs—of dropping flashcards. *Memory*, *16*(2), 125–136. https://doi.org/10.1080/09658210701763899

Lafleur, L. (2020). The indirect spaced repetition concept. *Vocabulary Learning and Instruction*, *9*(2), 9–16. https://doi.org/10.7820/vli.v09.2.lafleur

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

Leonard, L. B., Christ, S. L., Deevy, P., Karpicke, J. D., & Kueser, J. B. (2024). Retrieval practice and word learning by children with developmental language disorder: Does expanding retrieval provide additional benefit? *Journal of Speech, Language, and Hearing Research*, *67*(5), 1530–1547, https://doi.org/10.1044/2024_JSLHR-23-00528

Lin, C.-H., Zhou, K., Yang, S., & Sun, Z. (2023). Developing an evaluation framework for vocabulary-learning apps. *Interactive Learning Environments*, *31*(10), 7377–7391. https://doi.org/10.1080/10494820.2022.2068037

Lin, J.-J., & Lin, H. (2019). Mobile-assisted ESL/EFL vocabulary learning: A systematic review and meta-analysis. *Computer Assisted Language Learning*, *32*(8), 878–919. https://doi.org/10.1080/09588221.2018.1541359

Mishra, P. (2006). Affective feedback from computers and its effect on perceived ability and affect: A test of the computers as social actor hypothesis. *Journal of Educational Multimedia and Hypermedia*, *15*(1), 107–131.

Muqaibal, M. H., Kasprowicz, R., & Tissot, C. (2023). Evaluating the impact of spaced practice using computer-assisted language learning (CALL) on vocabulary learning in the classroom. *Language Teaching Research*. Advance online publication. https://doi.org/10.1177/13621688221146146

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. (2016). Effects of retrieval formats on second language vocabulary learning. *International Review of Applied Linguistics in Language Teaching*, *54*(3), 257–289. https://doi.org/10.1515/iral-2015-0022

Nakata, T. (2020). Learning words with flash cards and word cards. In S. Webb (Ed.), *Routledge handbook of vocabulary studies* (pp. 304–319). 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

Nakata, T., & Webb, S. (2016). Does studying vocabulary in smaller sets increase learning? The effects of part and whole learning on second language vocabulary acquisition. *Studies in Second Language Acquisition*, *38*(3), 523–552. https://doi.org/10.1017/S0272263115000236

Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed.). Cambridge University Press.

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Heinle Cengage Learning.

Nikoopour, J., & Kazemi, A. (2014). Vocabulary learning through digitized and non-digitized flashcards delivery. *Procedia - Social and Behavioral Sciences*, *98(6)*, 1366–1373. https://doi.org/10.1016/j.sbspro.2014.03.554

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

Pan, S. C., Dunlosky, J., Xu, K. M., & Ouwehand, K. (2024). Emerging and future directions in test-enhanced learning research. *Educational Psychology Review*, *36*(1), 20. https://doi.org/ 10.1007/s10648-024-09857-2

Papathanasiou, E. (2009). An investigation of two ways of presenting vocabulary. *ELT Journal*, *63*(4), 313–322. https://doi.org/10.1093/elt/ccp014

Pyc, M. A., & Rawson, K. A. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory and Cognition*, *35*(8), 1917–1927. https://doi.org/10.3758/bf03192925

Pyc, M. A., & Rawson, K. A. (2011). Costs and benefits of dropout schedules of test-restudy practice: Implications for student learning. *Applied Cognitive Psychology*, *25*(1), 87–95. https://doi.org/10.1002/acp.1646

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

Schmitt, N. (1997). Vocabulary learning strategies. N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 199–227). Cambridge University Press.

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

Serfaty, J, & Serrano, R. (2024). Lag effects for foreign language vocabulary learning through Quizlet. *Language Learning & Technology*, *28*(1), 1–21. https://hdl.handle.net/10125/73567

Simonnet, E., Loiseau, M., & Lavoué, É. (2025). A systematic literature review of technology-assisted vocabulary learning. *Journal of Computer Assisted Learning*, *41*(1), e13096. https://doi.org/10.1111/jcal.13096

Suzuki, Y., Nakata, T., & DeKeyser, R. (2020). Empirical feasibility of the desirable difficulty framework: Toward more systematic research on L2 practice for broader

pedagogical implications. *The Modern Language Journal*, *104*(1), 313–319. https://doi.org/10.1111/modl.12625

Terai, M., Yamashita, J., & Pasich, K. E. (2021). Effects of learning direction in retrieval practice on EFL vocabulary learning. *Studies in Second Language Acquisition*, *43*(5), 1116–1137. https://doi.org/10.1017/S0272263121000346

Teymouri, R. (2024). Recent developments in mobile-assisted vocabulary learning: A mini review of published studies focusing on digital flashcards. *Frontiers in Education*, *9*, 1496578. https://doi.org/10.3389/feduc.2024.1496578

Uchihara, T., Webb, S., Saito, K., & Trofimovich, P. (2022). The effects of talker variability and frequency of exposure on the acquisition of spoken word knowledge. *Studies in Second Language Acquisition*, *44*(2), 357–380. https://doi.org/10.1017/S0272263121000218

van den Broek, G. S., Gerritsen, S. L., Oomen, I. T., Velthoven, E., van Boxtel, F. H., Kester, L., & van Gog, T. (2023). Optimizing multiple-choice questions for retrieval practice: Delayed display of answer alternatives enhances vocabulary learning. *Journal of Educational Psychology*, *115*(8), 1087–1109. https://doi.org/10.1037/edu0000810

Webb, S. (2009). The effects of receptive and productive learning of word pairs on vocabulary knowledge. *RELC Journal*, *40*(3), 360–376. https://doi.org/10.1177/0033688209343854

Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford University Press.

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

Wilcox, A., & Medina, A. (2013). Effects of semantic and phonological clustering on L2 vocabulary acquisition among novice learners. *System*, *41*(4), 1056–1069. https://doi.org/10.1016/j.system.2013.10.012

Xodabande, I., Atai, M. R., & Hashemi, M. R. (2024). Exploring the effectiveness of mobile assisted learning with digital flashcards in enhancing long-term retention of technical vocabulary among university students. *Journal of Computers in Education*. Advance online publication. https://doi.org/10.1007/s40692-024-00347-6

Xueting Ye, S., & Shi, J. (2023). Investigating the potential of changing the smartphone system language to L2 for facilitating vocabulary learning and motivation. *Language Teaching Research*. Advance online publication. https://doi.org/10.1177/13621688221145565

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

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

Table 1
Flashcard Apps Evaluated

	Brainscape	*iKnow!*	*Memrise*	*mikan*	*Quizlet*	*WordHolic*
Version	4.20221201	5.3.7	2022.12.13	6.19.0	7.19.1	2.33.0
Developer	Brainscape	DMM.com LLC	Memrise	mikan Co., Ltd.	Quizlet Inc	Langholic Ltd.
Price	Free or 19.99 USD/month	1,510 yen/month	1,500 yen/month	Free or 600 yen/month	Free or 700 yen/month	Free or 120 yen/month
Primary target users	English speakers learning a wide range of subjects	Japanese speakers learning English	Speakers of diverse languages learning English, Spanish, Korean, French, German, Italian, Chinese, or Russian	Japanese speakers learning English	Speakers of diverse languages learning a wide range of subjects	Speakers of diverse languages learning a wide range of subjects
Examples of ready-made flashcards	Academic English, Spanish Level 0 to 4, French Level 0 to 4	TOEIC, Business English, English for studying abroad	Greetings, ordering at a restaurant, sightseeing, business	English vocabulary for college entrance exams, daily conversation, business English, studying abroad	N/A	N/A
Average user ratings (out of 5.0)						
App Store	4.7	4.7	4.8	4.8	4.7	4.7
Google Play	4.4	4.5	4.6	4.4	4.7	4.7

Approximate number of user ratings	17,000	15,000	1,610,000	98,000	465,000	54,000

Note. Ready-made flashcards include only those created by the app developers. N/A is given for *Examples of ready-made flashcards* for *Quizlet* and *WordHolic* because the developers of these apps do not provide ready-made flashcards.

Table 2

Comparison of Flashcard Apps

	Criteria	*Brainscape*	*Brainscape* (paid)[a]	*iKnow!*	*Memrise*	*mikan*	*mikan PRO* (paid)	*Quizlet*	*WordHolic*
Flashcard creation and editing	1. Flashcard creation	+	+	+	+	+	+	+	+
	2. Multilingual support	+	+	+	+	+	+	+	+
	3. Support for multiple words for target items and their meanings	+	+	+	+	+	+	+	+
	4. Support for flashcard sets	+	+	+	+	+	+	+	+
	5. Sharing flashcard sets	+	+	+	+	–	–	+	+
	6. Types of information for ready-made flashcards								
	L2 definitions[b]	–	–	–	–	–	–	N/A	N/A
	L1 translations	+	+	+	+	+	+	N/A	N/A
	Audios	+	+	+	+	+	+	N/A	N/A
	Phonetic symbols	–	–	+	–	+	+	N/A	N/A
	Images	–	–	+	–	–	–	N/A	N/A
	Videos	–	–	–	+	–	–	N/A	N/A
	Parts of speech	–	–	+	–	+	+	N/A	N/A
	Word parts	–	–	–	–	+	+	N/A	N/A
	Inflections	–	–	–	–	+	+	N/A	N/A
	Derivations	–	–	–	–	+	+	N/A	N/A
	Associations	–	–	–	–	+	+	N/A	N/A
	Example sentences or phrases	+	+	+	–	+	+	N/A	N/A
	User notes	–	–	–	–	–	–	N/A	N/A

	Criteria	*Brainscape*	*Brainscape* (paid)[a]	*iKnow!*	*Memrise*	*mikan*	*mikan PRO* (paid)	*Quizlet*	*WordHolic*
	7. Types of information for custom flashcards								
	L2 definitions	+	+	+	+	+	+	+	+
	L1 translations	+	+	+ +	+ +	+	+	+ +	+
	Audios	–	–	+ +	+	+ +	+ +	+ +	+ +
	Phonetic symbols	–	–	+ +	–	+ +	+ +	–	–
	Images	–	+	+	–	–	–	+	+
	Videos	–	–	–	–	–	–	–	–
	Parts of speech	–	–	+ +	–	+ +	+ +	–	–
	Word parts	–	–	–	–	+ +	+ +	–	–
	Inflections	–	–	–	–	+ +	+ +	–	–
	Derivations	–	–	–	–	+ +	+ +	–	–
	Associations	–	–	–	–	+ +	+ +	–	–
	Example sentences or phrases	–	–	+ +	–	+ +	+ +	–	–
	User notes	–	–	–	–	–	–	–	+
Learning	8. Presentation	+	+	+	+	+	+	+	+
	9. Retrieval	+	+	+	+	+	+	+	+
	10. Receptive recognition	–	–	+	+	+	+	+	–
	11. Productive recognition	–	–	+	+	–	+	+	–
	12. Receptive recall	–	+	+	+	–	+	+	+
	13. Productive recall	+	+	+	+	–	+	+	+
	14. Varied encounters and use	–	–	+	–	++	++	–	–
	15. Increasing exercise difficulty	–	+	+ +	+ +	–	+	+	+

	Criteria	*Brainscape*	*Brainscape* (paid)[a]	*iKnow!*	*Memrise*	*mikan*	*mikan PRO* (paid)	*Quizlet*	*WordHolic*
	16. Block size	–	–	+ 10, 25, 50, 100	–	+ increments of 10	+ increments of 10	++	++
	17. Interference avoidance	+	+	+	+	+	+	+	+
	18. Fluency development	–	–	+	+	+	+	+	+
	19. Automatic speech recognition	–	–	–	+	–	–	–	–
	20. Adaptive sequencing	+	+	+	+	+	+	–	–
	21. Retirement criterion	self-judgment	self-judgment	self-judgment, auto-drop	auto-drop	auto-drop	auto-drop	auto-drop 1	self-judgment
	22. Expanded rehearsal	–	–	+	+	–	–	–	+
	23. Feedback for learning	+	+	+	+	+	+	+	+
	24. Motivating feedback	–	–	+	+	+	+	+	–
Scores	Flashcard creation and editing	9	9.5	17	10	22	22	8	8
	Learning	6	7.5	13	13	9.5	13	11.5	8.5
	Total	15	17	30	23	31.5	35	19.5	16.5

Notes

[a] *iKnow!* and *Memrise* are available only for paid users. There are free and paid versions of *Quizlet* and *WordHolic* (Table 1). However, because the results of the analysis are the same for both versions of these two apps, they are not distinguished in Table 2.

[b] In ready-made flashcards for *Brainscape*, meanings of lexical items are all given in English, regardless of target languages. Consequently, although L2 definitions are available for English, they are not offered for other target languages.

Figure 1
Images Provided for Ready-Made Flashcards in iKnow

Figure 2
Videos Provided for Ready-Made Flashcards in Memrise

Figure 3
User Notes for Custom Flashcards in WordHolic

Figure 4
Link to a Dictionary Entry in mikan

Figure 5
Receptive Recognition, Productive Recognition, and Productive Recall Formats in iKnow

Figure 6
Example of a Ready-Made Flashcard Set That Involves Semantic Clustering

Figure 7
Automatic Speech Recognition in Memrise