MA Thesis

Flashcards and the *Memrise* App for

English Vocabulary Acquisition

2017

Woosong University

TESOL-MALL

Donald E. Hobbs

Flashcards and the *Memrise* App for

English Vocabulary Acquisition

Dr. David Kent

Submitted to the thesis committee

2017 December

Woosong University

TESOL-MALL

Donald E. Hobbs

Date of Signature

David Kent, Ph.D.
Thesis Advisor
Department of TESOL-MALL
Woosong University

December 29 2017

Eric Reynolds, Ph.D.
Committee Member
Department of TESOL-MALL
Woosong University

2017/12/15

Mark Love, Ph.D.
Committee Member
Department of TESOL-MALL
Woosong University

Dec. 15/2017

Abstract

    This study examines the effectiveness of two different personal study methods for the acquisition of English language vocabulary – paper flashcards and *Memrise*, a popular vocabulary learning mobile application (app). The study was conducted over 10 weeks at a private Korean information college (정보대) offering a variety of 2-year programs. Two culinary program classes of 19 students each participated in it. A repeated measure, quasi-experiment based on test scores before-and-after provided the quantitative data, along with student self-reports and tracking information from the app itself. Several qualitative measures were also used: questionnaires, focus groups, individual interviews, and finally, the personal experiences and observations of both myself and the other instructor. Results were positive for both study methods, although the paper flashcard users showed a significantly higher increase in test scores than those who used *Memrise*. The author makes several recommendations about the use of mobile flashcard apps outside of the classroom for increasing the likelihood that students will routinely and effectively use one to study vocabulary and also proposes ideas for future research hailing from the present study.

Donald E. Hobbs

Department of TESOL – MALL

Woosong University

Table of Contents

I.   Introduction

The popularity of mobile language applications (apps) utilizing spaced-repetition and flashcard-like features is at an all-time high and continues to increase. At the time of the present study, in October of 2017, the start-up fundraising website *Kickstarter* announced that its most successful app campaign in history just closed. It was a product called *Fluent Forever* (Wyner; Shulman, 2017). This proposed foreign language mobile app's methodology begins with learning through pronunciation, specifically utilizing minimal pair training and proceeds to spaced-repetition flashcards with pictures and sound. Its creator is a polyglot with no formal education related to foreign languages whose "third key" to fluency is that modern electronic flashcards (with spaced-repetition) "are a massive shortcut to memorization" (Wyner, 2017).

Flashcards have a long history for teaching and learning (Azabdaftari & Mozaheb, 2012). They have been a popular way to learn foreign language vocabulary for several decades. With the recent development of e-flashcard and vocabulary apps, learners who want to acquire English (L2) vocabulary have numerous product choices. Accordingly, the present study initially explores the history of flashcards and development of mobile apps that have led to the current market.

Purpose and Questions

Once this history has been explored, this study looks at vocabulary learning by EFL students in a single college setting using two different methods of study: paper flashcards and the popular *Memrise* app. The effectiveness of each method and the opinions of learners are analyzed. The hope is to inform educators so that they can decide whether using these methods is appropriate in their own settings. If it is, the information in the present study should assist them in designing a vocabulary study program that is worthwhile and indeed, more effective than the one carried out here. This research is therefore important to classroom instructors.

A pilot study similar to the current one was carried out in South Korea in 2013 for 3 weeks. In it, the *Memrise* web-based application was made optional for students to study 60 vocabulary items taken directly from the course book. The study found that many students either did not use it at all or very little. The researchers concluded that a similar study conducted over the course of a semester and with larger class sizes might yield more robust results and suggested that the *Memrise* mobile-based app would make for a worthwhile future comparative study (Kent & Sherman, 2013). Although it was neither the inspiration for nor aim of this study to take up where theirs left off, the present study does to some degree take the researchers up on their suggestions.

The present study also touches on a number of themes relevant to using programs or apps for vocabulary acquisition: changing technology and its impact on the availability, cost, and features of popular market options; various spaced repetition schemes; and self-study by students, among others. These are all topics of interest not only for educators, but for L2 language learners themselves. At its core then, the present study is a simple comparison of the effectiveness of paper flashcards versus a mobile vocabulary learning app, which hopes to answer 4 key questions:

1. How effective are paper flashcards or *Memrise* for learning English vocabulary?

2. How do popular apps compare to each other in terms of features and performance?

3. How are both methods of study used by language learners and instructors?

4. How can such use – by students and instructors alike – be improved upon to make it more effective?

II. Background

History of Flashcards

Flashcards have a long history for both teaching and learning purposes that goes back to the 1800s. Various online sources (Quizlet, Wikipedia, Flashcardsguru.wordpress.com) credit Mortimer as the first teacher who was known to use flashcards, which were included in a book in 1834. Her phonics cards were designed for L1 children and lacked the "flip" feature of modern L2 flashcards. It seems likely that at least a few teachers, missionaries, or anthropologists used paper flashcards to study a foreign language in the 1800s, although I could find no record of it online or in academic literature. Regardless, widespread self-use of them by adults as a popular foreign language tool did not occur until the advent of modern scientific research and marketing. The fields of cognitive psychology and applied linguistics both played key roles through the introduction and application of "spaced-repetition" to adult L2 language learning.

Spaced Repetition

The idea that learning spaced out over time is more effective than if it takes place over a short single "cram" session is attributed to Ebbinghaus (1885). Cognitive psychologists call this the "spacing effect" and it has been proven in numerous studies since. In fact, "the spacing effect is one of the oldest and best documented phenomena in the history of learning and memory research"

(Bahrick & Hall, 2005). When applied to vocabulary study, it means that a stack of 20 vocabulary flashcards divided into even smaller stacks with optimal spacing will be remembered better compared to studying all 20 words at one go (Kornell, 2009).

In the late 1960s, several cognitive psychologists (Melton, 1967; Bjork, 1970; Landauer, 1970) also studied changing the timing of repetition as a way to improve short-term memory recall. A scholar of applied linguistics applied this concept to language learning, introducing what he called "graduated-interval recall" (Pimsleur, 1967). According to the Pimsleur website "if learners were reminded of new words at gradually increasing intervals, each time they would remember longer than the time before. He documented the optimal spacing for information to move from short-term into long-term, or permanent, memory." (www.pimsleur.com/the-pimsleur-method.) His research was used to develop and market the Pimsleur foreign language courses, which are still among the most popular examples of quality audiolingual courses on the market today (Choe, 2016, pp. 108-9).

Another pioneer in language acquisition was Lietner who introduced a method of flashcard study which was later marketed as the "Leitner system." It relied on the principle of spaced-repetition involving 5 stages/decks of cards (Leitner, 1972). Among current flashcard websites and apps, *Cram, MemoryLifter,*

and *VocabularyTrainer* advertise that they use modified versions of the Leitner system.

The concept of language being transferred from short-term to long-term memory through increasing intervals of time between subsequent review of previously learned material was not only the basis of both Pimsleur courses and the Leitner method, but remains the basis of the spaced-repetition software (SRS) in today's vocabulary mobile apps, including *Memrise*.

CALL for L2 Vocabulary Study

The 1980s. In the late 1980s and with the gradual spread of personal computers, the first CALL software utilizing spaced-repetition appeared. One example was *SuperMemo*, a vocabulary learning program created and released in 1987 by Wozniak (Jones, 2010). For me, it is the best example of early foreign language learning CALL software not only because it was one of the earliest, but because it is still a popular choice in today's competitive market as both an online and mobile app. Additionally, the 1987 spaced repetition algorithm "SM-2" was made public and served as the basis of later algorithms used by many of today's popular apps including *Anki*, *Tiny Cards* (by *Duolingo*), and *Memrise* (BlueRaja, 2016).

1990 – 2010. In the 1990s, with the advent of the internet and increasing popularity of foreign language learning, an increasing number of CALL programs

aimed at foreign vocabulary acquisition were produced. Most of the popular ones made use of spaced repetition; however, very few were ever mass-marketed and thus did not become popular with the average L2 (adult) learner. The most notable exception was the *Rosetta Stone* CALL software first released in 1992, which is now on Version 5 and among the most popular software for learning a foreign language. (Duffy, 2017). Like most CALL programs, it is much more robust than basic flashcard apps or flashcard-like apps such as *Memrise*, and includes "learning, writing, and grammar exercises" as well as the ability to use a microphone to have learner pronunciation evaluated (www.rosettastone.eu). The most common criticisms of it have always been its high cost and the inadequacy or limitations of using pictures and words in the target language as the basis for "learning" exercises (www.pcmag.com, www.brighthubeducation.com, www.economist.com).

Mobile Flashcard Apps

SuperMemo (initial release 1987) and *Mnemosyne* (2003) were both available online long before being developed into mobile apps in 2010. *SuperMemo* was also available as an app on the "pocket PC" before iPhone in 2010, then Android in 2012. Many of today's popular vocabulary apps appeared online when what could properly be called smart phones also made their appearance, 2006 – 2010. This includes *Anki, Brainscape, Quizlet, and*

StudyBlue. Of these, *Anki* and *StudyBlue* released a mobile app during this time; the other two were released as web-based flashcards and only developed a mobile app later: *Brainscape* in 2010, and *Quizlet* not until 2012. *Memrise* launched its website in private beta mode in 2009, in public in 2012, and its mobile app in 2013.

Today's market.   Despite the fact that vocabulary mobile apps are really a very recent development in foreign language learning, today's learner has a large number of them to choose from. These include *Anki, Brainscape, Chegg (Flashcards+), Cram, Course Hero, Duolingo (Tiny Cards), iKnow!, MemoryLifter, Mnemosyne, OpenCards, Quizlet, StudyBlue, SuperMemo, Tiny Classroom, Vocabulary Trainer,* and the one used in the present study, *Memrise*. A majority of these are now available to both Android and iPhone users for free. Several are free to download, but include Pro versions or pay "in-app" features. Some, like StudyBlue, which was originally "pay" ($10 on iPhone in 2009), have since become free (Phillips, 2009). A blog on the *MemoryLifter* website in 2012 sums up the market situation: "A new version of *MemoryLifter* has been released. The popular flashcard software is now open source and all learning content that was developed and formerly sold by *LearnLift* is available to the community for free" (http://www.memorylifter.com/news.html). The trend over the last several years, driven by market competition, has been for

website-based apps/programs to become free and also to be released as mobile apps available on both of the main mobile operating systems (iOS and Android), with most features free or inexpensive. (See Appendix A.)

In 2013, the founder and CEO of *Transparent Language, Inc*. which offers a variety of products for both the U.S. military and the public (through *Babbel*) commented on the current market. It had become a "bottom-feeding" one, where few consumers were willing to pay much for foreign language material with so much available for free, even though a majority of it was not necessarily high quality and usually geared toward beginning learners (M. Quinlan, December, 2012).

Within such a market, learners can not only use flashcard apps for free, sharing and creating their own decks, but they can also find a large quantity of website-based material (ranging from poor to fair quality) for free as well. This material includes major features of the language, the alphabet, basic pronunciation, common phrases, dictionaries, and vocabulary lessons, depending on the particular language and website host. Many such sites have been created by individual learners themselves, but large language companies like *Babbel, Transparent Language, Tandem, Busuu, Linzapp,* and *DuoLingo* them too, if only to advertise their better quality pay courses. (For a typical example, go to: www.babbel.com and select a language.) At the high end of the market, are

courses that have withstood the test of time and can command a relatively high price, either because of their quality or because they are supported by large customers like the U.S. government or military. These include *Pimsleur* and *Rosetta Stone* courses, along with various CALL style courses from the list of above companies (Duffy, 2016).

Flashcard Effectiveness

Flashcards have been used for vocabulary study for a long time because they are easy to use and presumed to be effective. Surprisingly, however, few studies have tested how effective they are compared to using word lists, notebooks, or dictionaries. As one study stated, "although the use of flashcards for language learning dates back many years, the number of research studies done in this area is limited" (p. 50). As these researchers also pointed out, the ones that exist mainly involve children – not adults trying to learn a foreign language (Azabdaftari & Maheb, 2012). In 1993, two researchers in the Netherlands extolled the advantages of flashcards in terms of memory, the spacing effect, and the Leitner system. They concluded that "the memorizing of words with the hand computer [the Leitner system] is clearly superior to the memorizing of words by means of inflexible lists, in whatever form." And, "even without a repetition system working with word cards has great advantages" (Mondria & Mondria-De Vries, 1993, p. 56). However, no comparative experiment involving flashcards or lists

was conducted. Another study suggested flashcards were of limited advantage over lists and showed only little difference in vocabulary gains, but with better long-term retention (Nakata, 2008). The effectiveness of flashcards is largely assumed, and may depend on factors such as the individual learner to a large degree. It should not be surprising that using flashcards seems to be beneficial and popular for many language learners, but this does not mean that they are the most effective – or even an ideal – way to acquire L2 vocabulary. Numerous researchers point out the limited benefits of learning vocabulary absent context (Krashen, 1989; Parry 1991; Oxford & Scarcella, 1994).

Oxford and Scarcella have divided L2 acquisition into three categories: *decontextualized, partially contextualized,* and *fully contextualized*. Under their framework, since no context is provided, flashcards are a "decontextualized" method for vocabulary learning. Using them is simpler than using the partially or fully contextualized tools and methods, and is another reason they have remained popular, even with the wide availability of more sophisticated English language mobile apps over the last two decades.

In theory, *Memrise* might qualify as partially contextualized according to Oxford and Scarcella, depending on how definitions are entered; for example, by using sentence completion plus an image. Practically, however, extremely short definitions (in L1 or L2) are better because of how the app utilizes multiple choice,

displaying up to 6 definitions at one time for a user to choose from is awkward and takes up too much screen space. A timer that runs down also favors this type of input because it takes too much time for the user to make his/her way through more contextualized material. Thus, the features of the *Memrise* app itself ensure it is most likely to be used as a "decontextualized" method of study, the same as other flashcard apps on the market.

Features of Flashcards

Comparing word cards to contextualized learning, Nation and Waring (1997) stated:

It [learning from context] is by far the most important vocabulary learning strategy and an essential part of any vocabulary learning program. For fast vocabulary expansion, however, it is not sufficient by itself. There is no research that shows that learning from context provides better results than learning from word cards . . . Direct learning of vocabulary from cards is a kind of form focused instruction which can have the same benefits, perhaps even more markedly so, as form focused grammar instruction . . . There are other advantages for using word cards. They can give a sense of progress, and a sense of achievement, particularly if numerical targets are set and met. They are readily portable and can be used in idle moments in or out of class either for learning new words or revising old ones. They are specifically made to suit particular learners and their

needs and are thus self motivating (Nation & Waring, 1997, "How much

vocabulary and how should it be learned?," para. 5).

Several positive features of flashcards: portability, motivational nature,

and their being made by the learners themselves, all make flashcards – paper or

electronic – ideal for studying foreign language vocabulary. First, whether given

a word list or making one him/herself, the learner is in charge of choosing which

words he/she needs to learn or review. This is obviously true of paper flashcards,

but all of the flashcard-like apps on the market today also allow users to import

lists/make and use their own cards. Flashcards are also a learner-centered and

self-directed method of study. Thus, they fulfill the important role of self-study

(Oxford, 2011).

Flashcards provide immediate feedback (once the card is flipped) without

risk of public failure, followed by self-correction and reflection, which are all key

elements in language learning. The range of strategies used by language learners

differs greatly (Gu & Johnson; Lawson & Hogben, 1996). From my own

experience, I know this applies to flashcards. They are very versatile in the way

they can be/are used by different learners. Discussing flashcards as 1 of 3

"decontextualized" methods of vocabulary study (alongside word lists and

dictionaries), Oxford and Crookall (1990) stated:

With creativity and thought on the part of the learner, flashcards can

indeed be employed to provide greater context. For instance, learners can sort flashcards into piles representing different groups of words (e.g.,nouns, adjectives, adverbs, verbs, prepositions; words that are already learned vs. words still needing to be learned; past tense forms vs. present tense forms). Learners can tape flashcards to particular objects like lamp, table, and chair signified by the words on the cards, thus providing a visual (and to a degree tactile) context. Visual context might also be added by taping relevant pictures to the cards. Learners can arrange flashcards on the floor in a kind of semantic map, with related words closer together and unrelated words farther apart. Finally, students can add context by writing the new words in complete, meaningful sentences on the cards (p. 12).

From my own experience, I know these to all be common uses, but there are other possibilities as well. Learners typically labeled as "effective" or those who display a high "self-efficacy" for vocabulary study (Mizumoto, 2012), often use the reflective break offered after flipping the card to mentally employ additional techniques such as mnemonics, word association, or to construct their own sentences with the word, which qualify as Vocabulary Learning Strategies, or VLS (Gu & Johnson, 1996; Lawson & Hogben, 1996; Mizumoto, 2012).

Finally, flashcard use automatically requires the learner to review words. While most learners will probably not discover on their own any method as complex or effective as either Pimsleur or Leitner, again, from my own

14

experience (and as Oxford and Crookall infer), effective paper flashcard users often instinctively or through trial-and-error end up creating separate decks or stacks of words and reviewing ones that they do not know or have trouble remembering more than ones they find easy. They often discover the utility of the spacing effect and periodic review of different decks based on how well they know the words, which is fundamentally a form of "spaced repetition."

With modern flashcard apps, however, SRS does away with much of the work described above and the distinction between "effective" and "ineffective" learners by automatically deciding the optimal review of words based on when the learner used the app last and his/her own personal history of missed words. Once many dedicated (effective) learners become acquainted with the various spacing schemes/algorithms used by different e-flashcard and vocabulary learning apps, they are capable of forming their own opinions about the comparative effectiveness of each for their own learning (4tops, 2011; Pleco, 2017).

Mobile Apps vs. Traditional Methods

Does this mean that vocabulary apps in general are more effective than paper flashcards or other traditional methods of vocabulary study? Most studies say yes. A recent meta-analysis concluded that "mobile devices have moderate effect sizes on both ways of vocabulary learning (productive and receptive)" and a "moderate effect on all aspects of vocabulary learning" (Mahdi, 2017).

15

A number of studies related to the effectiveness of mobile apps for foreign language learning have been conducted in recent years, but 7 were found to be relevant to the present one.

Three studies were carried out in 2012. The first of these took place in Taiwan. There were 32 participants and 3 groups: a control group who relied on their own study methods (16), a group who used a vocabulary app "Learning King" developed in Taiwan on their tablet PCs (8), and a group who used "Learning King" on their mobile phones. This app had a "number of features, such as word definition and usages, pronunciation practices, a personalized vocabulary organizer, and a self-quizzing function" (p. 1057). The tablet PC group significantly outperformed the other 2 (with M = 91, SD = 9) on a final test. The control group using traditional "paper" methods had a much lower score (M = 79.4 / SD = 7.5). The students using mobile phones showed a much larger variance (SD = 15.5), with an average score of 78.5, which was actually slightly lower than the group using traditional methods (presumed to be word lists and vocabulary notebooks, etc.). However, when asked about their perceptions, most students thought that the use of the app increased their performance and said they would use it in the future (Hung & Chao, 2012).

The other 2 studies conducted in 2012 compared paper flashcards with flashcard apps, similar to the present study. The first was by Azabdaftari and

Mozaheb. Its focus was whether vocabulary would be better learned using flashcards or a similar e-vocabulary program like *Memrise*. The researchers chose a mobile-based vocabulary one (unspecified and presumed to be non-commercial) that utilized Spaced Repetition Software (SRS) technology. 80 students at a university in Tehran participated, and at the end of seven weeks, a vocabulary test and interviews were conducted. On the independent test, the mean scores of the group using Mobile-learning (M-learning) were statistically higher (65/100) than that of the control group using flashcards (45/100). The study also had a qualitative aspect, and based on both the discussions of questions given to participants and test scores, concluded that the mobile vocabulary program was significantly more effective than regular paper flashcards (Azabdaftari & Mozaheb, 2012).

Another study conducted in 2012 was at Southwestern Oklahoma State University by Burgess and Murray and focused on paper and flashcard apps, both separately and together, but for L1/L1 psychological vocabulary. Regardless, it has been included here due to its unique findings: namely, that flashcard use declined when students were only provided with an app, and that as an optional outside-class study method, most students did not use the app at all. Students preferred the paper flashcards, but when provided with both, they were much more likely to use the app. Reasons given by students for not using the app

included concerns over battery life and not surprisingly, being distracted by other apps, texts, etc. that come with M-learning. This study importantly suggested that paper flashcards could still be useful for many learners.

Nikoopour and Kazemi sought to answer how effective all 3 types of flashcards (paper, online, and mobile-based) were in a 2014 study. 115 participants were divided evenly into 3 groups, each utilizing one of these types as their method of study for 10 weeks. The study was the most rigorous of those reviewed. It made use of the same 70-item pre-and-post-test, 4 progress tests as a motivation tool, a 30-item questionnaire (Likert 5 scaled), and interviews. Importantly, the same flashcard content was provided to all 3 groups (nobody created their own cards or lists). This content was apparently in English on both sides, and included a definition, example(s), synonym(s), and antonym(s). A special website was designed for the online group and a special software app created for the mobile group. The study found that on the post-test, both the paper and mobile flashcard groups significantly outperformed the online one. Statistically, there was no significant difference between the mobile and paper group. The researchers concluded that the portability aspect of both the paper flashcards and mobile app was the key factor in the performance differences. The learners preferred using mobile phones to study vocabulary (Nikoopour & Kazemi, 2014).

The last 2 studies reviewed were both undertaken more recently, in 2015. The first was at Bangkok University and similar to previous ones, completed over a relatively short time period of 7 weeks. It compared a control group of 40 students, who did vocabulary paper exercises in class, with a study group of 40 who did them on their mobile phones, via SMS. Vocabulary mean scores of students in the paper-based group and SMS group were 22.85 and 22.57 (out of 50 points) and increased to 29.70 and 33.25 respectively after the intervention. From t-test analysis, the post-test mean scores were significantly higher than the pre-test mean scores in <u>both</u> groups. Perhaps more importantly, students reported that they preferred the vocabulary delivered via mobile phones in all cases (Suwantarathip & Orawiwatnakul, 2015).

The second study conducted in 2015 was at Jiujiang University in mainland China, over 55 days. It used an English vocabulary mobile app "Word Learning CET6" especially designed by the author herself for Chinese students. The study compared the scores of students who used the app versus a control group left to their own notebooks and word lists. On a pre-test of 100 items selected from a master vocabulary list of 1274 items (the CET6 test) the mean scores were 35.68 for the control group and 35.13 for the test group. On the post-test, the control group's average score was 43.56, while the test group's was 50.52, a percentage difference of less than 7%. (Wu, 2015).

None of the studies reviewed used a popular web-based or mobile app. Perhaps, as Nikoopour & Kazemi (2014) stated, "finding a mobile software compatible with all types of cellphone operating systems was impossible" (p. 1369). Nevertheless, I knew from my own experience that finding a suitable one was possible. To do so, the differences between commercially available mobile vocabulary apps needed to be examined.

Differences in Flashcard Apps?

Various technology websites, such as *App Advice* and *Make Use Of* offer useful reviews for language learners. However, a comprehensive and scholarly comparative study was also carried out in 2011, by Nakata. In it, he used seventeen criteria to assess nine popular "robust" flashcard (CALL) programs: *LearnThatWord*, *Word Engine*, *WordChamp*, *P-Study System*, *MemoryLifter*, *VTrain*, *Quizlet*, *SuperMemo*, and *iKnow!* (Nakata, 2011). Several have already been mentioned. The first 3 in his study, however, are now either out of business or lowly rated and unpopular as mobile apps on Google Play and App Store. The fourth, P-Study, seems to only be available for learning Japanese. (This highlights the role of the current online competitive public market in determining which apps become or remain successful.) The remaining 5 programs, which are all successful as apps, were all considered as potential choice for the mobile app used for the present study.

Nakata found variety in the way the 9 flashcard programs were designed, but concluded that most were designed to maximize effectiveness. Seven of them, for example, allowed users to create their own cards, and add contexts, audios or images. *Iknow!* was rated the best overall, but all of them had some shortcomings and room for improvement. (2011). Notably, none of them was "designed to encourage generative use of words." For example, the word should be "used in different senses, collocations, inflections, grammatical functions, or sentence patterns" when it is reviewed (p. 26). It is interesting to speculate that the previously mentioned features of *Memrise* that would allow for partially-contextualized content, versus the definite decontextualized nature of most of the other popular flashcard apps might similarly allow for the generative use of words Nakata found lacking; however, the way in which the app is typically used does not. Thus, the lack of this feature is still a shortcoming of all of today's popular flashcard app programs.

The programs in Nakata's study were considered robust programs because they offered the benefit of being able to upload audio, video, or photos as part of the word definition. However, the line between "simple" flashcard apps and robust programs has become blurred or erased since Nakata's study, as even simple flashcard apps on the market now allow for audio and image upload.

21

Importance of spaced-repetition. One of the reasons typically credited for positive results achieved by flashcard apps or similar vocabulary CALL programs is the use of spaced repetition. Miles and Kwon (2008) quoted Nation's 2001 study as stating that one promise of CALL was its ability to deliver spaced-repetition, a "key principle" for vocabulary learning (p.1). Today, nearly every flashcard app on the market advertises SRS. Its benefits for foreign language vocabulary learning have been frequently studied, but Miles and Kwon (2008) aptly stated:

Studies in memory research (Baddeley, 1990) and vocabulary learning (Bloom and Shuell, 1981; Dempster, 1987) have confirmed that the student in our example following the spaced repetition method would have much better retention of the words than the other student who followed the massed repetition method, even though the total amount of time on task was equal for each learner. This phenomenon is based on the principle that periodic reviews keep new information from being forgotten. With each review, the information is strengthened in the memory (p. 2).

In other words, as discussed earlier, the "spacing effect" does work. This fact, together with mobile technology and a competitive market, are the reasons why today's vocabulary apps use SRS.

Best App?

Most of the mobile flashcard apps on the market today are similar overall, just as the CALL programs in Nakata's 2011 study were. Since his study, SRS has since become a standard feature. Similar to the other studies reviewed, Nakata's did not present a clear choice for "best app" for use in the present study simply because of changing technology over the last several years.

Personal experience. Accordingly, I relied heavily on my personal experience in my selection of an app suitable for the present study. Since 1987 I have often used paper flashcards to study vocabulary for various languages. In 2011, I began to replace individual language-specific e-dictionaries and stacks of paper flashcards with mobile apps, including 2 vocabulary apps previously mentioned: *Anki* and *Tiny Classroom*, which I have used extensively for learning Indonesian and Thai. I have tried several of the other apps in Appendix A, including *Memrise*. In 2016, I used *Anki* and *Chegg* flashcard apps with students for a TOEIC course.

Suitability. I not only relied on my personal experience, but on that of other professional teachers, many of whom recommended *Memrise* as an option. While it seemed suitable, I wanted to be sure. I decided on 5 critical criteria the app for the present study must have:

1. Popular in today's market on both Android and iOS

2. Free to download and use without upgrades (via Google Play and the
   App-Store).

3. Easy for new users to create an account and start using it.

4. Easy for a teacher to create a list of words/deck/course and share it with all
   students.

5. Utilizes spaced-repetition (as a non-pay feature).

The former editors of the TESOL CALL Interest Section Software List stated that there is no one-size-fits-all or magic pill when it comes to selecting software. Their software checklist included the important item: "How much money do you have to spend? Maybe this should be the first question." (Healy, 1997, "First Steps"). Criteria numbers 1 and 2 are similar, and the reasons for their inclusion clear (Nakata, 2011). Unlike the previous studies examined, the present one could not develop its own app if it were to carry out one its main purposes: to be useful to both instructors and learners alike. A popular, available, and free app was the only choice suited to this purpose. Similar to many institutions, ours would not pay for students to download a vocabulary app, and the participants – students or instructors – were not willing to pay either. As already stated, the expectation in today's market place is that a worthwhile flashcard-like app should be free.

Criteria 3 and 4 are also related. In my experience, nothing quite demotivates learners – or instructors – from using a new technology more than having difficulties figuring out how to use it. None of the participants in the study were especially tech-savvy and the instructors could not afford to take up much class time addressing technical issues. I knew, for example, that using the *Anki* app to import lists or adjust spaced repetition was too complicated for the students in the study. The mobile app needed to be easy for everyone to quickly start using. Such considerations all fall under the categories of "Learner Fit" and "Teacher Fit" as discussed by Hubbard in his CALL Evaluation Framework (Hubbard, 2015).

Many of the apps under consideration (Appendix A) failed to meet criteria 1 and 4. *Memrise* met all 5. Satisfied with *Memrise* as a choice, I also decided to evaluate the app to provide useful information for other instructors or learners. The evaluation checklist used (Appendix B) was a simple one I created, but based partially on Son's checklist and recommendations for evaluating websites (Son & O'Neill, 2005). The importance of criterion 5 – (SRS) has already been discussed.

Differences in spaced-repetition. Evaluating *Memrise* required doing some research about its technical specs compared to those of apps with which I was more familiar. Different apps use different SRS, and thus different algorithms. These control the intervals between review. Because the SM-2 was one of the earliest and most successful spaced-repetition algorithms and was made available

to the public, it makes sense that a number of company's technical teams have used it as a starting point for their own modified spaced-repetition scheme. The algorithm itself does not warrant discussion here, but has been included at the end of Appendix A.

Although I did not receive an answer to inquiries from *Memrise* regarding its algorithm (most likely do to proprietary concerns), a number of technical blogs claim the spacing review is based on SM-2, similar to other apps. But, what makes it unique? Several features, actually. It uses "plant" and "water" sessions initially. The plant session is actually designed to be an initial learning session, with a frequent amount of review. The assumption here is that this is the first time a learner is being exposed to any of the vocabulary words contained in a particular online course. After the plant session, it proceeds to a "water" session, which is meant to be a true review session, with less frequent repetition of words. After that, it follows a fairly set schedule based on a 180-day program. In other words, overall, this schedule was designed based on a learner whose goal was to finish and remember all the words in a course, or multiple courses (600 words plus), long-term, with regular periodic study spaced out up to 6 months. After the plant and water sessions, the frequency with which words are reviewed becomes less and less frequent. The long-term review scheme for *Memrise* has been reported on a company blog (Memrise Community Hub, 2016) and is included in Appendix B.

Since *Memrise* uses a schedule based on the assumption of a long-term study of foreign language vocabulary, spaced over fairly even intervals, an obvious question arose: given such a design, would it work well for acquiring vocabulary over the short-term, for which many instructors and learners actually use it? Additionally, it is a one-algorithm-fits-all design (as are all the apps except for *Anki*). Would it address the needs of either particularly quick or slow vocabulary learners? However, despite any potential weaknesses, among the many flashcard and vocabulary apps considered, *Memrise* had the most advantages: it was free, available on both Android and iOS, was slightly more robust than a simple flashcard app, and was easy to sign up for, download, and use. Additionally, it provided a way to track student use, progress, and difficulty for the present study.

Increasing Learners' Use of a Mobile App

Obviously and regardless of which mobile app is used to acquire vocabulary, SRS does not work if students do not use it. Personal study habits outside the classroom are therefore an extremely important factor for efficiency of any flashcard app. One strategy to increase the likelihood of students using mobile devices for vocabulary study outside class was proposed and tested fifteen years ago by Thornton and Houser at Kinjo Gakuin University in Tokyo. They used "push" media, delivered to students' phones on a set weekly schedule. Some

vocabulary study websites and apps today offer this option. In their study, this option proved to be quite successful although the study did not address long-term use, but rather, took place over fifteen weeks (presumably a semester). It should be noted as well that the vocabulary "pushed" to students included mini-stories, which qualify as fully contextualized material, in addition to mere definitions or L1/L2 pairs utilized in simple flashcard and flashcard apps. (2001). Several more recent studies (Lu, 2008; Hayati et al., 2013; Wu, 2015) suggest that sending content or study reminder messages via SMS to students on a regular schedule causes more of them to study. The latest study at Jiujiang University was similar because the English vocabulary mobile app (CET6), like *Memrise* in the present study, was installed on their phones and could be used anywhere, anytime. Students did not have to wait for content delivery, but were simply sent a text reminding them to use their app. The researcher did not quantify the "push" effect, but commented that "she was told" by participants in the test group that they "sometimes started to learn via CET6 immediately after they received a 'pushing' message at the time of idling," which the control group did not do (Wu, 2015). It is not a miracle solution in other words, but it is better than not reminding students.

Conclusions of Research To-Date

While numerous studies have concluded that mobile flashcard apps or similar CALL programs seem to be more effective than paper flashcards,

vocabulary exercises, or notebooks, most were not particularly rigorous in eliminating or controlling one or more factors that may have affected the outcome. One was the students' English existing proficiency. In some studies, tests did not take account of how many words students in each group already knew. The ways in which learners might have used either flashcards or the apps in question on their own was ignored. Finally, varying amounts of time that individual users spent using the flashcards and apps were not carefully recorded in many studies.

As a result, one might ask whether the use of mobile vocabulary apps for vocabulary study outside of class is at all worthwhile. The Oklahoma study found that many students did not often use such apps on their own when they had free time, and at intervals that according to research, would provide the most benefit in tandem with SRS (Burgess & Murray, 2013). Wu (2015) reported similar dismal study times for students using a mobile vocabulary app.

If the promise of flashcard apps is indeed that learners will use them smartly outside classrooms to study and reinforce vocabulary and take control of their own personal learning, the studies on Multimedia-Assisted Language Learning (MALL) to date are either silent on this issue, or have concluded that university students, at least, are unlikely to do so unless they are prompted to via push technology such as that utilized in the Thornton and Houser study (2001).

III.   Methods

The present study is conducted using mixed methods similar to the 2014 studies by both Kent and Sherman, and Nikoopour and Kazemi. Quantitative data shows how effective each method of vocabulary study was comparatively. Questionnaires and focus groups are used to examine student motivations and preferences about each method of study and also help verify the accuracy of the data. A variety of information helps inform instructors who may want to use either of these methods or conduct a similar study in the future.

Setting

National setting. In South Korea, as in other countries, there is a hierarchy of universities and colleges, both public and private. Unlike many countries, South Korea has a declining student population which will continue in the future. All universities and colleges are aware of this trend and have felt its impact. Most have made efforts to increase the number of foreign students, mainly from China, they are admitting into their universities and colleges.

In 2015, the Education Ministry began ranking all universities and colleges nationwide and placing them into 5 designations, A – E, which determine the rate and manner in which they will have to lower their enrollments, with severe penalties for schools in the bottom D/E categories (Jung, 2015). Within this setting, the ranking of a college by the government becomes critical; schools in

the top 3 tiers are likely to succeed or at least remain viable and open in the future; those in the bottom tiers (D-E) have an uncertain future and may be forced to close. Within this system, the information college where the present study is conducted can be considered a typical third tier school; it is a well-known regional one that is less selective and easier to be accepted into than top-tiered universities. The quality of the student body reflects this, including their overall lower proficiency in English compared to Korean students at many other universities.

       Study setting. Both groups of students in the present study are enrolled in a two-year program in which culinary courses are taught mostly in English by foreign chefs. Both groups attend courses together in the same building on campus, and the English courses they take are identical and best described as being mixed ESP (cooking) and general English. The culinary program also helps its interested, motivated, and higher achieving students apply for and receive internships working abroad as chefs during/at the end of their second year. Among the English language professors who teach at the information college, the students in this program are generally considered to have a higher level of English than most of their fellow students in other programs. Future goals that may include either an internship working in a restaurant overseas or a career as a chef are a motivating factor for some of the students in the program when it comes to learning relevant English cooking vocabulary like that tested in this quasi-experiment.

Participants

Instructors. Two instructors participated in the study. I was the first, and my class was designated group A for the study. The second instructor's class was designated group B.

Table 1

Instructor	G	Age	Teaching Experience	Method
A	M	40s	12 yrs − EFL, different schools, various Asian countries	Flashcards
B	M	40s	12 yrs − EFL, different schools in South Korea	Memrise

Learners. Similar to their instructors, the students in groups A and B were extremely similar. Both were two classes of first-year students, ages 19 − 21 from the same college in South Korea, in the same culinary department. The two classes were also equal in size, with 19 students each. Both groups had a similar spread of English proficiency as indicated by a university simulated TOEIC speaking test administered in June, 2017, which both instructors felt was generally an accurate indication of relative English proficiency given their previous experience teaching these students. The only notable difference was gender: there were more males in group B. However, one limitation of the present study − and every study reviewed − was that the aptitudes of the learners in each group were not measured, and could not be compared.

Table 2

Group	Gender	TOEIC Band (no. of students)	Group	Gender	TOEIC Band (no. of students)
A Flashcards	16 F 3 M	7(1), 6(2), 5(7), 4(5), 3(4)	B Memrise	7 F 12 M	8(1), 7(1), 6(3), 5(6), 4(5), 3(1), 2(2)

Data Collection

Research design. The present study took place from the beginning of the 2017 fall semester until mid-November, over a total of 11 weeks. The first step necessary was the creation of a master vocabulary list for the students to learn for fall semester. The list was limited to cooking words. I focused on items that the students should know by the end of their first year, but felt might not, based on class the prior semester. Glossaries found in the text series *Career Paths: Cooking* levels A2 and B1, provided a handy source of words yet unstudied. This three-level series text is one my colleagues and I had taught single chapters from in the past, but only parts of the 'A1' and 'A2' texts. I omitted words that I felt most students already knew, and about 150 representing dishes/foods (nouns), cooking techniques and types of food (adjectives), and cooking methods (verbs) were placed on the list.

Next, I added a good number of culturally specific food words not in the text such as "chicken fried steak" or "Lamington," which may or may not have

come up in the previous semester. Other words relevant to Korean cuisine were added as well: 'date,' 'silk worm pupae,' and 'mortar and pestle,' for example. The list was shown to the instructor B. He suggested the addition of half a dozen words ('faux caviar,' 'deconstructed'). Finally, a Master Vocabulary List of 233 words for the students to know at semester's end – and for the study – was complete (Appendix C). Based on user blogs, I estimated that in *Memrise*, 1 word = approximately 1 minute to "learn" so that 100 words = 100 minutes, then doubled that time to estimate the maximum time needed by slowest students, accounting for difficult missed words and mandatory review. I estimated that despite the vocabulary list being fairly long, the Group B learners should all be able to learn it all by using *Memrise* less than 1 hour per week.

Test design. Given two nearly identical instructors and groups, a single vocabulary list for the semester, and a core research question on the relative effectiveness of paper flashcards versus *Memrise*, a two-group, pre-test, post-test design was an obvious choice. It was a simple but effective way to measure the effectiveness of the two study methods. As Aron, et al. (2003) have pointed out, such repeated measure designs have a lot of power, but also disadvantages from the point of view of results because of alternative explanations (factors) for the difference in scores (p. 26). In the present study, because of the striking similarity in the two groups' instructors and students, I felt strongly that such variables could

34

be controlled through coordination and conscientious observation. A standard multiple choice test of 61 words (1 extra due to a numbering mistake) with choices A – D was created and administered week one of the quasi-experiment. During the test design, care was taken to select words that: 1) a majority of students would be unlikely to know; 2) were relatively important and frequently appear in a food context. The same vocabulary test was administered week 11 of the semester, at the end of the study.

Reported known words. In addition to test scores, other related data were collected: each student self-reported how many words from the entire list he/she knew at the beginning and end of the study. To do this, each student simply counted and reported the number of words he/she knew from the Master Vocabulary List.

Intervention Design

Set-up. Once the initial vocabulary test had been administered and the word list emailed to students in the study, it was time to fully explain the vocabulary objective and to introduce each group to the study method they would be using. Group A learners were told they would be using paper flashcards; group B learners were told they would be using the *Memrise* app. This was accomplished at the beginning of week 2. For both groups, the instructors explained that the goal for the class was for every student to know all the words

on the Master Vocabulary List at semester's end, and that a second 'random' vocabulary test would be given in week 11.

Paper flashcards. As instructor A for Group A, I first played a video showing how to use flashcards for vocabulary study creating one stack of 'words I know' and a second stack of 'I don't know' words (https://www.youtube.com/watch?v=MfecA4K4nVg). Next, I demonstrated the principle using a stack of Korean vocabulary flashcards (one side Korean; the other English) and also showed how to periodically review words, maintaining two separate stacks (one "know" and one "don't know").

Students were all given 40 blank flashcards. They accessed the Master Vocabulary List on their mobile phones to create their own deck and were explicitly reminded not to include any words that they already knew. They all wrote the English word on the front side of the card and were allowed absolute discretion as to what they put on the back. All of them then used a dictionary app or the internet to look up the unknown word's meaning and any pics. Many chose to write mostly a Korean definition on the back side, while many wrote a short English definition. Many also used both English and Korean. The only directions they were given was to use any simple definition that would help them remember it the best and to be able to choose the correct English definition on a second multiple choice (post) test.

During week 8, I reminded Group A students to use the 2-stack method based on my observation that many students had reported either little or no flashcard use outside of class and that during the short amount of in-class time given to vocabulary study, some did not seem to be using the cards as they had been shown.

Memrise **group.** First, a new *Memrise* course for group B students, named "Cooking Vocabulary 2017" was created. Next, 213 of the words from the Master Vocabulary List were entered into it. 20 words were either not critical or already known by most students. This was checked against the pre-test. A course with 20 fewer items than the vocabulary list would help offset the fact that the flashcard group (A) had the inherent advantage of not studying items they already knew. Most of the L2 definitions in "Cooking Vocabulary 2017" were in simple English, with images for 78 of them and more than one image for several. Korean was included together with English for 22 of the L2 definitions; for 18 words the L2 definitions were solely in Korean (words like "vinegar, tofu, walnut") with images.

Next, instructor B had each student in group B go to the *Memrise* home page to sign up for an account if he/she did not already have (one student had already used *Memrise* in another course). Then, the students all downloaded the mobile app onto their phones from either Google Play Store or the App-Store. Finally, they all used their phones to join the course. After confirming all of them

had done so and knew how to use *Memrise*, Instructor B gave no further technical instructions for the remainder of the semester.

Study motivators. Students in both groups were verbally reminded by their instructors to study vocabulary on a weekly basis. Two emails were also sent out to each group during the 10 week study, one during week 3 and another during week 7. Instructors A and B decided not to implement stronger measures such as "pushing" weekly goals or sending SMS reminders to students – as in the Thornton and Houser (2001) or Wu (2015) studies. Both felt that verbal reminders would suffice. Two additional motivators convinced the instructors that most students would complete the vocabulary goal: 1) most of their homework grade (10% of the total course grade as set by the institution) would be based on improvement in vocabulary test scores and overall study time that could be verified; and 2) food rewards, initiated by the students themselves: 1 candy bar for the first 40 words learned and verified, then another for every 20 word thereafter. Both groups were promised a paid-for pizza party if the group's average score on the final post-test was 45 (out of 61).

Study Time

Outside class. For both groups, each student self-reported the amount of time spent using either paper flashcards or *Memrise* outside of class. This was collected and recorded periodically directly by both instructors, starting at the end

of week 3 (see Appendix D).

In-class study time. The amount of time spent in class by students using their respective study method was limited, controlled, and coordinated between the 2 instructors so that the time allotted for study was nearly the same for both. Both instructors kept a record of this time.

Some of the words on the vocabulary word list were contained in the units of the *Career Paths: Cooking* text. Both instructors coordinated the text units they taught in order to ensure limited and identical in-class exposure to the vocabulary items through reading and discussion, or what Oxford and Scarcella call "fully contextualized." Specifically, both taught the same 5 units from the text in-class. There were only 4 vocabulary words from these units that appeared on the 61-item vocubulary test.

Questionnaires

An initial questionnaire with items related to students' experience with their respective study method and their impressions of it, along with their English study habits and motivation, was also administered and collected at the end of the first week of the study. Four questions on the initial questionnaire were omitted from the final one, which was administered week 11, at the end of the study. All other questions were asked again in order to see if there was any difference in either group's study habits or attitudes about their respective method of study.

The final questionnaire had an additional 2 questions not on the initial one (see Appendix E).

Focus Groups

At the end of the study, all the students from both group A and B were interviewed, in groups of 3-4 at one time. These focus groups were semi-structured, with open-ended responses and dialogue between members of the group encouraged. I conducted them in English, with Korean used as necessary with any students who were not comfortable or fully able to express themselves well in English (see Appendix F).

The groups were organized according to their general English proficiency (based on both TOEIC band scores and class performance) and overall increase in test scores. The flashcard group was divided into 4 groups: 1-low English proficiency but a high (126 – 176%) increase in test score; 2-high proficiency + high increase (109 – 143%) in test score;   3-mixed proficiency + mixed increase (34 – 95%) in test score; and 4-high English proficiency but a disappointing increase (27 – 58%) in test score. The groups for the *Memrise* students were the same, but with the addition of a group 5, comprised of lower proficiency students who also performed poorly on the final test (-10 – 21%). These groups are more clearly illustrated in the following table:

Table 3

Focus Group	Members	English Proficiency	Score Increase (No / %)
A	Flashcard learners		
A-1	17A, 16A, 15A, 11A	Low / 3-4 TOEIC	High: +19-30 / +126 − 176%
A-2	7A, 14A, 1A, 19A, 10A	High (class performance)	High: + 25-31 / +109 − 143%
A-3	3A, 6A, 12A, 13A, 18A	Mixed (3-6 TOEIC)	Mixed: +18-26 / +34 − 93%
A-4	2A, 4A, 5A, 8A,13A	High (5-6 TOEIC)	Low: +10-18/ +27 − 58%
B	Memrise learners		
B-1	14B, 18B, 19B	Low (2-4 TOEIC)	High: +20-25 / +88 − 156%
B-2	12B, 13B, 15B, 16B	Low (4 TOEIC)	Med: +12-20 / +54 − 78%
B-3	3B, 4B, 6B, 7B	High (5-6 TOEIC)	Med: +15-21 / +55 − 78%
B-4	1B, 2B, 5B, 8B	High (5-8 TOEIC)	Low: +10-17 / +36 − 55%
B-5	10B, 11B, 17B	Mixed (3-5 TOEIC)	Low: -3-6 / -10 − 21%

The thoroughness and variety of methods used ensured that the present study was as rigorous as it could practically be given the institutional setting and standard ethical considerations. Participants were notified that all data collected would: 1) only be seen by Instructors A and B initially; 2) be kept completely anonymous, and 3) not impact them in any way besides the homework and rewards aspects already mentioned. There were a number of benefits to students.

In addition to exposing them to a vocabulary study method/tool they could use in the future, the present study provided a worthwhile learning objective (a set vocabulary goal of content words) for the semester. No other homework was given to either group during the semester, so that the students were free to study vocabulary (or not). As hoped, the various methods employed provided interesting results, which are reported in the following section.

Results

In this section, quantitative data consisting of scores on the vocabulary test, students' self-reported data, and *Memrise*-generated statistics are reported first. Next, I report the data I have labeled as qualitative; namely, the results from the questionnaires and study groups. How both types of data answer the research questions posed at the beginning of the paper has been left to the Discussion section.

Quantitative Results

The quantitative data analyzed consisted of the quasi-experiment (each group's initial and final scores on the vocabulary test), instructor reported data, student reported and *Memrise*-generated data. Appendix D shows initial and final vocabulary scores for each student in groups A and B.

There was a dramatic improvement in scores for group A (flashcards). The mean score on the initial test was 24.26, while the average on the post-test was 45.95. Initial test scores ranged from a low of 12 to a high of 37 (out of 61), with a variance of 41.9 and a standard deviation of 6.47. Final test scores ranged from a low of 31, dramatically higher than the initial low score, to a high of 54 – similarly a marked increase, with a variance of 45 and a standard deviation of 6.7. This shows a somewhat uniform increase in scores, across the entire group, which was an impressive 97 % on average. The mean of the reported study times was

5.38 hours, with a median and mode of 3. There was not a great variance in reported study times, although there were 2 outlying students who reported studying 24 hours each!

A look at group B, which used the *Memrise* app, showed a similar pattern: The mean score for group B pre-*Memrise* was 26, only slightly higher than the mean for group A, and the average post-study score was 41, which was not as great an increase as group A. Scores on the initial test for group B ranged from a low of 17/61 to a high of 39/61, showing a similar spread to group A, with a variance of 42.88 and a standard deviation of 6.55, nearly identical to those of group A. Final test scores for group B/*Memrise* ranged from a low of 25 to a high of 55, again similar to those in group A, but with a much larger variance of 72 and standard deviation of 8.5. On average, group B showed a 64 % increase in scores. The reported study times of group B were similar to group A, with a mean of 4.67, a median of 4, and mode of 3, with only slightly more variance but unlike group A, no extreme study times (the highest was 9 hours – by the student who showed the greatest increase in test scores).

In short, group A (top) showed a more uniform and greater increase in test scores than group B (bottom) given approximately the same study time, as shown in Figure 1.

Figure 1
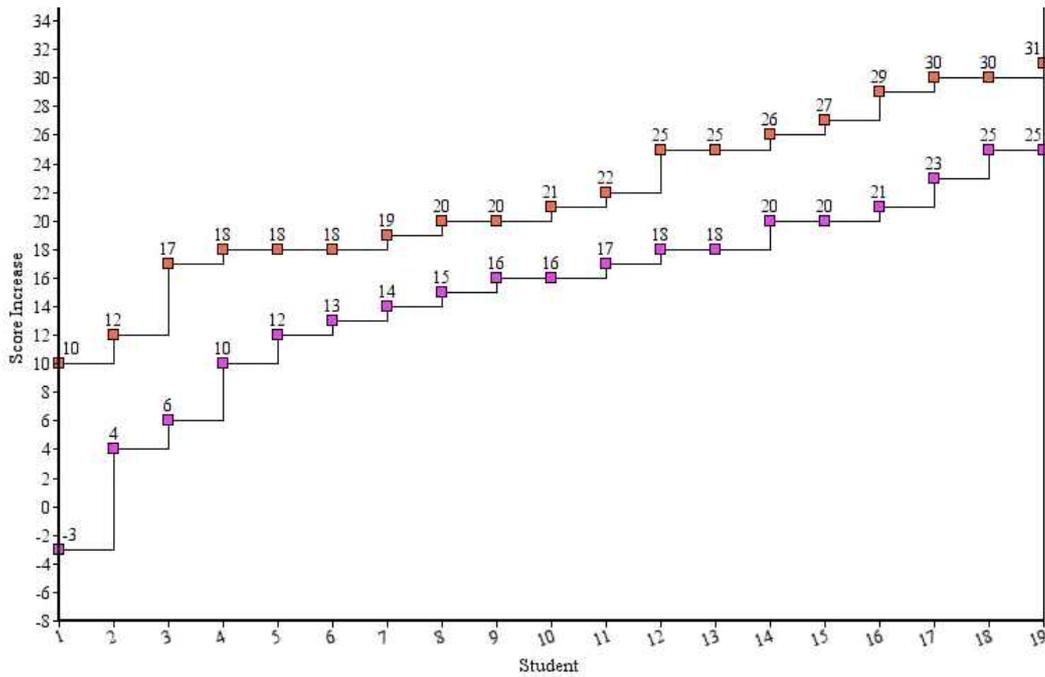
Existing proficiency as factor? As noted previously, students were assigned anonymous number plus letter designations based on their relative overall proficiency in English (as determined by TOEIC band scores) so that student 1A had a band score of 7 and a high proficiency, and student 19A a band score of only 3 and relatively low proficiency. This lent itself to determining whether there was any significant correlation between students' existing proficiency in English and their improvement in test scores for either method of study. As shown from the scatter diagrams (Figures 2 & 3) there did not seem to

be any noticeable correlation. To confirm this, correlation was also tested using the Pearson calculation method. For group A (flashcards), there was technically a weak negative correlation, r = -.0307 and for group B (Memrise), r = - 0.343. Given the range of vocabulary scores for each TOEIC band (existing level of proficiency), however, it is safe to say there was no correlation.

Figure 2

Group A

Eng. Proficiency & Score Increase

Figure 3

Group B
Eng. Proficiency & Score Increase

47
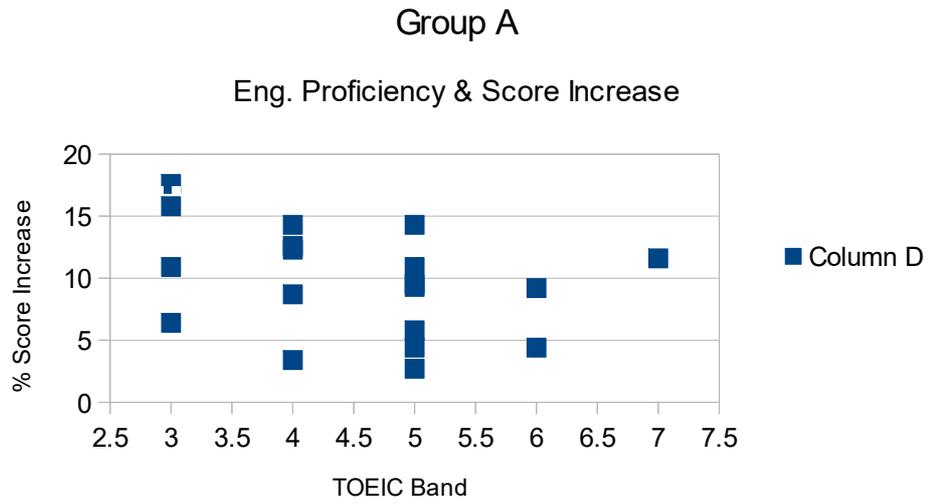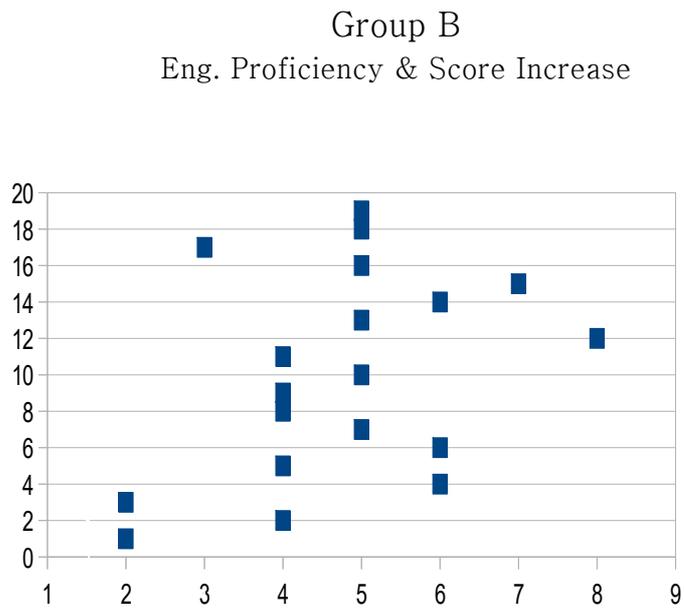
T-test. There is no population for this quasi-experiment because it used a sample of convenience (as in most EFL studies) rather than a random one. As such, an independent samples t-test was performed to compare the difference between the two groups' vocabulary gains, using the following calculations (two-tailed distribution):

$$s^2_p = ((df1/(df1 + df2)) * s^2_1) + ((df2/(df2 + df2)) * s^2_2) =$$

$$((18/36) * 36.22) + ((18/36) * 52.09) = 44.16$$

$$s^2_{M1} = s^2_p /N1 = 44.16/19 = 2.32 \quad / \quad s^2_{M2} = s^2_p /N2 = 44.16/19 = 2.32$$

$$t = (M1 - M2)/\sqrt{(s^2_{M1} + s^2_{M2})} = 6.74/\sqrt{4.65}$$

t = 3.12,   p = .0035. The result is significant at both $p < .05$ and $< .01$.

Effect size. One of the most common ways to compute a standardized effect size is using a measure known as Cohen's *d.* It is the appropriate effect size measure for two independent samples that have a similar standard deviation and are approximately the same size, which is true in this study. It measures the chance that a person picked at random from one group will have a higher score than a person picked at random from the other (probability of superiority). It is determined by calculating the mean difference between the 2 groups, or ($M2$ - $M1$), then dividing the result by the *pooled* standard deviation, where *SD* pooled =

$$\sqrt{((SD_1^{2} + SD_2^{2}) / 2)}$$

Group A: M = 45.95,   SD = 6.71 Sample Size = 19

Group B: M = 41.37,   SD = 8.49, Sample Size = 19

(41.37 - 45.95) / 7.651934.   Cohen's d = 0.6, which is a medium effect size.

With small sample sizes (below 20) however, the effect size actually has an upward bias of up to 4%. To correct for this, a modified Hedge's g, is often performed:

d x (N − 3)/N − 2.25) x √N − 2/N, or g = .599 x .955 x .946 = **.54,** which is still a medium effect size, but slightly smaller than 6.

Reported Data

Instructor-reported. In the Methods section, it was reported that instructors for both groups closely coordinated and kept track of in-class study time. Here is the class log from Instructor A (paper flashcards):

Table 3

Date	Minutes	Group A (Flashcards) Instructor Notes
9/5	15	
9/13	15	3 students playing on phones, not studying!
9/20	15	25% of class did not use time to study. 1st candy bar for 40 words successfully learned (quizzed on-the-spot) awarded!
9/26	0	2nd 40-word winner!

9/27	0	3rd winner!
9/28	15	4 students = doubtful they studied.
10/15	15	7 students now have learned 40 new words; 2 students another 20 words.
11/08	20	3 students not studying.

Instructor B reported weekly in-class study time via Google docs. At the end of the study, he estimated a similar total of 90 – 110 minutes total in-class study time had been given to group B (*Memrise*) students, with similar misuse of the time by only several.

Student-reported. Students reported the amount of time spent <u>outside</u> of class using either method of vocabulary study. This was collected verbally by the instructors at the same intervals for both groups, and is shown in Appendix D.

Memrise **Data**

Direct data from *Memrise* is displayed as two tables in Appendix G. The first table's data was generated on November 1. Notice that some of the data is for the previous month; some of it is cumulative. 6 students are not represented; possibly this was because they had not accessed the course at all during the 30 days prior; although this could also be due to technical issues. The most noticeable figure in Table 1 is the dismal study time, which shows that only 2 students spent more than an hour using *Memrise* in October.

The app failed to spark enough interest for most students to continue using it outside of class as the semester progressed. The cumulative total number of words learned up to November was slightly more positive. Many of the students had used the app in September in order to learn a sufficient number of words to receive an initial reward (candy) and to feel confident they had learned the words A – E sufficiently well to do well on the quiz (which in fact they did).

The second table covers achievement from the first table until the end of the course, with a period of overlap for study time 18 – 31 October. Some data for 1B, 2B, 4B, 13B, 16B, and 18B was not retrievable, for unknown reasons. It appeared on the "leader board stats," but not in the official group ones. This remains an unexplained glitch which did not negatively impact the study.

Memrise course completion. Overall, only 3 students completed the entire *Memrise* course. Another 4 finished at least 50% of it. The data suggests most managed to void the advantages of the spacing effect by using the app the first two weeks and again before the final test, but little during the time between. There seemed to be a correlation between the percentage of the course completed and final test scores, as indicated in Figure 4.

Figure 4
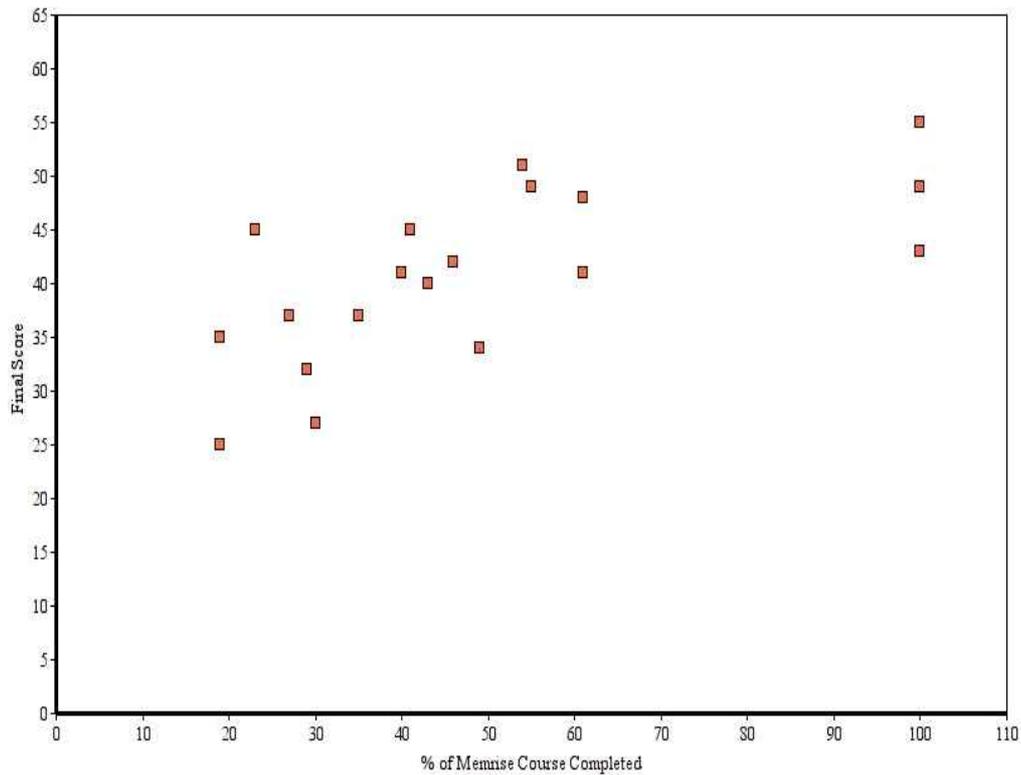
The precise correlation coefficient was determined using the Pearson calculation method:

X Values: $\sum = 932$ Mean $= 49.053$ $\sum(X - M_x)^2 = SS_x = 12258.947$

Y Values: $\sum = 786$ Mean $= 41.368$ $\sum(Y - M_y)^2 = SS_y = 1298.421$

X and Y: $N = 19$, $\sum(X - M_x)(Y - M_y) = 3032.632$

R Calculation: $r = \sum((X - M_y)(Y - M_x)) / \sqrt{((SS_x)(SS_y))} / r = 3032.632 / \sqrt{((12258.947)(1298.421))}$

52

r = 0.7601, which is a strong positive correlation.

Correlation between the percentage of the *Memrise* course completed and the raw <u>increase </u>in final test score was similarly calculated, with result r = .37 (a weak correlation).

Qualitative Data

Now that the quantitative results have been analyzed, it is time to turn our attention to the qualitative data, which came from asking the participants themselves what they thought. This was accomplished through student questionnaires and focus groups, and the observations of the 2 instructors.

Questionnaires. The initial and final questionnaires were administered to both groups as discussed in the Methods section of this paper, and results tallied. As mentioned, it was hoped that some useful information could be gleamed from students' responses, and it was. For analysis, the questions were divided by purpose into 3 categories: 1) those aimed at determining students <u>experience</u> with either method of study or with English language apps in general; 2) those related to students <u>motivation</u>; and 3) those related to their <u>opinions</u> about their respective methods of study (flashcards or *Memrise*).

What were the similarities between the two groups' experiences, motivations, and opinions? Regarding experience, many participants in group A indicated that they had used flashcards before (I found this to be inaccurate based

on further questioning) and also had used an English vocabulary app or game on their phones. In contrast, the students in group B reported no previous experience with *Memrise* and only about half indicated they had used a vocabulary app on their phones.

Looking at motivation, both groups agreed on several statements: they would definitely work with foreign chefs in the future, and most thought they would go on an internship abroad to work their 2nd year in the program. So, they felt they needed to know English vocabulary. They also usually looked up unknown English words in a dictionary. However, they were not motivated enough to use an app to study English. It is interesting to note that they saw the need for English vocabulary study, but until the experiment, had not really done much of it.

Finally, looking at the initial opinions of each group toward their method of study shortly after being exposed to it, more students in group B felt that *Memrise* was a good way to learn vocabulary quickly, was more effective than lists, and was fun, compared to group A who were more ambivalent towards flashcards and overwhelmingly chose 3 – "maybe" for their response to these same questions. This should not be surprising given the amount of time the average Korean college student spends using his/her smart phone; one would expect students to have a more favorable opinion of an app than paper cards.

At the end of the study, a comparison of initial and final questionnaires was made. For the most part, there were no changes in students' behavior in regards to either using a dictionary to look up unknown English words or using an English vocabulary game or app on their phones. For the flashcard card group, however, there was a slight shift in regards to dictionary use (from "fairly often" to "often"). This may have been due to the fact that they had to use their dictionaries (on their phones) extensively in order to make their flashcards and write definitions for the notebook method of study a majority also made use of.

Although their behaviors changed little, both groups' opinions about their respective study method did shift after 10 weeks of use, to being more positive about its effectiveness, and liking it more. For group B, their responses to "I like the way *Memrise* works" and it "is more effective than memorizing words from a list" went from somewhat ambivalent (maybe) to a general agreement (Yes, I think so). A notable exception to more positive sentiment was in regard to the statement "it seems like a good way to learn vocabulary quickly." Students' attitudes remained about the same, and only 2 students were convinced that this statement was 100% true on the final questionnaire. This was probably because a learner's opinion about *Memrise* has already been formed after only several hours of use, and is unlikely to change with more.

For the flashcard group, there was a noticeable shift in their responses to the same statement: "it seems like a good way to learn vocabulary quickly." Most agreed with it at study's end. And, similar to the *Memrise* group, flashcard students thought flashcards were more effective and easier to use at the end of the study than at the beginning.

While feelings about their method being "fun to use" were mixed, ranging from ambivalent to positive across both groups, more flashcard learners favorably changed their opinion after prolonged use, while *Memrise* learners did not, probably for the reason already suggested (an opinion about the app is formed early on after little use, while it takes a longer time using flashcards to form a definite opinion).

Focus Groups. As planned, small focus groups were conducted at the end of the study. The focus groups were recorded, analyzed, and reported on (Appendix F). Special attention was paid to individuals who exemplified the characteristics of each group. The approximate number of hours total that students reported studying was used for the "total est. study time" field in Appendix D. Student responses to their study patterns provided rich data for the Discussion and Conclusion sections of this paper.

Much of the data was unexpected, and one disclosure by students in particular called into question the nature of the quantitative data collection itself,

changing the very methods being compared. Most students in group A did not actually use flashcards to study outside of class. Rather, about 1/3 relied on a mixed flashcard and notebook/word list method, and over half exclusively the latter. Some attempted to use flashcards to memorize the first 40 words, but lost them and then switched to a notebook method (6 students or nearly 1/3 of the class lost all or some of their cards). In fact, only 2/19 students seem to have used cards exclusively/most of the time outside of class. With almost 100 hours of flashcard use in-class, the entire group could now best be labeled as "learners using a mixed flashcard and notebook method" versus one using flashcards.

Notebook method. The notebook method group A students chose was nearly the same: they looked at the Master Vocabulary List, which was emailed to them on their mobile phones, or in a few cases, printed the list. Then, for words they did not know, they looked them up in a dictionary online and wrote the word with definition in a notebook, thereby creating a word list which they used to memorize words from. Students reported that they "felt more comfortable" using that method, it was "a habit" or that they were familiar with it. A majority seemed to feel that it was a more efficient/better way to memorize vocabulary, although a few disagreed. One particularly astute student commented "I was taught this way since middle school." 2 students reported saying the words and definitions aloud while they wrote them.

Of the 2 students from group A who relied on flashcards outside of class rather than a notebook, one (17A) showed the most improvement, and the other (11A) was also a top performer, but only after a reported 10 hours of study time with them.

Summary of Results

The qualitative data showed significant improvement in vocabulary test scores for both groups A and B, indicating that both study methods were effective. Comparatively, group A's score's showed a greater and more uniform increase. Most of the group A learners switched from flashcards to a notebook as their primary method of studying vocabulary outside of class. Since nearly all of group A used a mixed flashcard plus notebook study system, it is impossible to draw conclusions about the efficiency of paper flashcards alone versus the *Memrise* app. The fact remains, however, that the learners using a mixed flashcard and notebook method outperformed their counterparts in the present study, most likely because they were able to learn a larger portion of the words on the Master Vocabulary List compared to the students who used the *Memrise* app for a similar number of study hours. Another possible factor in their superior performance is the process of looking up each word in a dictionary then writing it out along with a definition – required for both flashcards and notebook methods. We can only speculate which of these helped group A remember the words better than group B.

V. Discussion

Overall, the results of the present study replicated those of Hung & Chao (2012), which were previously discussed: the group using traditional methods slightly outperformed the one using a mobile app, but the app was preferred and thought to be more effective. Together, the quantitative and qualitative data helped to answer the original research questions posed in the Introduction.

Question 1. How effective are paper or mobile app flashcards for learning English vocabulary? The data clearly showed that both methods were effective, but that group A using paper flashcards and notebooks was more effective in this case. Why? The *Memrise* data and self-reports show that group A, on average, finished studying a larger number (percentage) of the vocabulary words than group B did: 167 (self-reported), versus 112 (self-reported) by *Memrise* students. This number is corroborated by the official *Memrise* data, which shows that on average, group B students learned 105 words, thus completing only 49% of the 213-word course. The approximate 6-7 hours average total that both groups spent in and outside class studying vocabulary allowed Group A to study more words. Using flashcards and notebooks allowed these learners to focus solely on words that they did not know because they made their own cards (or lists), substantially decreasing the total number of target words. Additionally, if a word was learned quickly, they moved it into "known" status and gave it little further attention.

In contrast, *Memrise* required all learners to start with the letter A and progress through all 213 of the words in the course, encountering many they already knew along the way. The repetitive nature of *Memrise* does not allow a learner to flag or identify a word he/she knows to quickly move it into a "100% known" status; even words a learner gets correct are repeated to some degree, and certainly more often than a fast learner might choose to do so on his own.

Question 2. How do apps compare to each other in terms of features and performance? As discussed earlier and shown in Appendix A, most flashcard-like apps on the market today have similar features and are free on both major mobile operating systems. Among these, *Memrise* is atypical because it does not incorporate the basic 'flip' feature which displays a definition. Rather, it uses the course words' definitions as multiple choice items and occasionally also makes the learner write the word itself for a given definition when it is presented to him/her.

The qualitative results confirmed those of the studies discussed previously in Section II, "Mobile Apps vs. Traditional Methods" (Azabdaftari & Mozaheb, 2012; Hung & Chao, 2012; Nikoopour & Kazemi, 2014; Suwantarathip & Orawiwatnakul, 2015). The participants preferred using the mobile app. Specifically, Group B learners liked using *Memrise* and thought it was effective. They would choose to use it again over notebooks or another method of their own. These learners did not, however, compare the app to others or evaluate any

60

technical features in detail. The qualitative student data regarding the *Memrise* app was added to my own to complete the evaluation of it (Appendix B). Overall the app rated a 2 on a scale of 1 – 3, with a mix of positive and negative features, as one might expect from any of the popular flashcard apps.

Question 3. How are both paper flashcards and vocabulary apps used by language learners and instructors alike? The data showed that with *Memrise*, most students initially used it frequently, but as the novelty wore off, used it less to the point that some did not meet vocabulary goals set for the course. With paper flashcards only a few students used them outside class frequently at first. A majority started using them in earnest the week prior to the test. For both groups, the problems were the same: learners did not study enough to learn all of the vocabulary items chosen for. They also did not use either method throughout the semester regularly, and therefore, did not take advantage of the spacing effect in order to maximize vocabulary retention.

Question 4. How can use – by students and instructors alike – be improved upon to make it more effective? One possible suggestion; namely sending reminders via SMS has already been discussed. The data collected from the present study itself did not directly answer this question, but reflection on it did lead to the recommendations made in the following Conclusion section.

VI. Conclusion

In the present study, a group of Korean college students using a mixed flashcard and notebook method of study outperformed a similar group of peers using the *Memrise* app. After overall hours of study where accounted for, the difference between the two groups was significant but explicable. Left to freely study after initial instruction, very few learners in either group effectively used the method they were given. Neither study method motivated the majority of its students to spend sufficient time outside of class to complete a course goal of learning 233 words in 10 weeks, although those using cards and notebooks came much closer to reaching it. Learners in the flashcard group fell back on a method of study using notebooks, which was more familiar and comfortable for them. For both groups, few learners took advantage of the "spacing effect." Rather, a majority did most of their study at the beginning of the semester and even more so at the end, cramming for the vocabulary test during the week prior.

Advantages and Disadvantages

Both study methods were found to have advantages and disadvantages, but students overwhelmingly preferred using the *Memrise* app; they felt it was a handy, effective way to learn vocabulary, and importantly, stated they would use it in the future if required to learn vocabulary, versus a notebook or another method.

The disadvantages of *Memrise* have carefully been noted in Appendix B; but in reality, most are minor or able to be compensated for by following the upcoming recommendations in the present study. In fact, there were few negative comments. One student who did not like the app did not like any mobile app. Only a few students thought that the ads present in the free version were a distraction. And, only a few thought (like I did) that the frequency of repetition during the "planting" phase was too much. The biggest disadvantage with paper flashcards is how younger learners view them: they might conclude that they are effective versus other methods, but since they are not readily available, easy to make or carry around, and are easy to loose, they probably will not embrace them. Most flashcard users in the present study, in contrast to *Memrise,* abandoned cards in favor of keeping a notebook, and stated that they would prefer either using a notebook or try a vocabulary app in the future. And, when all is said and done, is it not the learners' opinions that matter? Personally, I will continue to use both paper and flashcard apps, but for my younger students, paper flashcards are simply not a favorable option.

On the other hand, the advantages of *Memrise* are apparent. First, it is a mobile app. Unlike a stack of paper cards, it is not likely to be left or forgotten somewhere since it is already on the learner's phone and always with him/her. Almost everyone in the present study remarked on this. Another advantage is its

spaced repetition algorithm, which makes learners repeat some words more than others – without thinking about it. Some students admitted that if left to their own willpower; however, they could not use a mobile app efficiently since they would be tempted to use their phones for social purposes rather than study. This insight is in fact a dilemma without obvious solutions.

For instructors who want to track student study time and difficulties, *Memrise* has a "stats" button along with "leader board" stats allows the course manager to find and track information on every member of the group: study time for the last 30 days, cumulative number of words learned, the number of words in "long term memory," the number of "difficult" or oft repeated words along with a list of these same words, the percentage of the entire course completed, and the number of points cumulative and during the last 30 days.

Overall, *Memrise* is a practical option for short-term (one semester) vocabulary study, provided the recommendations for use are considered and followed.

Recommendations for Use

The implementation of these 2 methods for outside-class vocabulary study were only partially successful in the present study. How could it be improved? For instructors who want their students to use paper flashcards, *Memrise*, or another similar vocabulary mobile app outside class for self-study,

there are several considerations:

The unavailability of paper flashcards in many places. The most successful student in Group A mentioned this as a disadvantage. Proper pre-made note cards are not typically sold in stationary stores in many countries in like South Korea. Learners who want to make their own have to go to a print shop, find a suitable thickness of paper, and cut their own, or, use thinner lined cards. Some students will lose paper cards and not carry them around.

The features of any mobile app are slightly different. I recommend using similar criteria to that found in the *Choosing an App* section of this paper, as well as examining and applying a checklist like the one in Appendix B to any app before using it for an entire class of learners.

It is likely that many students will not be motivated to use any such study method regularly in a manner that is most effective.

Lack of motivation. This last point is critical and a major failure of the present study. Both instructors reminded their students to study vocabulary at home every week, but it had only a negligible effect – if any – on whether they did or not. Although they seemed relatively motivated to learn cooking vocabulary (many commented on the usefulness of the Master Vocabulary List), many still did not study much outside of class. Several studies cited in the literature suffered from similar problems. The reasons seem to be the same:

lack of time, being a low priority, being too busy or lazy, forgetting to do it, and so on. The carrot/stick approach used in the present study (candy for the number of words known plus a minor overall grade penalty for not knowing them) did not sufficiently motivate students. The average hours of outside-class study for both groups was a disappointing 4-5 hours over a 10-week period.

Both instructors in the present study felt that the time and effort spent on both study methods was worthwhile because of the positive effect each had on students, who recognized the need to learn the words from the vocabulary list. In the case of *Memrise* learners, many continued using it in order to complete the *Cooking Vocabulary 2017* course, even after the present study was concluded. By implementing procedures that encourage, monitor, and ensure regular use, both instructors are confident that *Memrise* can be a tool for successful outside-class vocabulary study. Instructors who wish to use it should:

1. Explain to students the benefits of spaced review, and how spaced repetition for studying with programs like *Memrise* works.

2. Consider allowing students to make their own courses in *Memrise*. The act of using a dictionary to choose a correct definition from context, inputting it, and uploading images or sound will provide an initial preview of all words and partially contextualized exposure to them. It will also prevent students from spending time "learning" words they already know. One practical way to do this

would be to divide the students into small groups based on the approximate number of words they know from testing/self-reporting. Each group could divide up the input of words among its members. Once its members had completed its own course it might choose another group's or use another vocabulary list to make another course. This would be more complicated than just giving all the students the same instructor-made course, but more beneficial.

3. Make sure the vocabulary list for making any *Memrise* course is not too large! Estimate the amount of time it will take. A quick and easy formula is: 5 min/day = 3 words, 20 words/week, 80 words/month; or, 100 words = 200 minutes to learn; then add a bit more time for slow learners.

4. Have students choose (by clicking on the icon) a "daily goal" adequate for the number of words they need to learn. For most students this will be a mere "5 minutes/day." Share the weekly "leader board" stats in class weekly and reward winners in order to promote competition and motivate students.

5. Have all students "follow" each other so they can see the number of points other students have earned.

6. Keep track of stats on a weekly basis and privately encourage students who are lagging to step-up their performance.

Another possibility for increasing the likelihood that students will study regularly for short time periods is to use a so-called "push" system already

discussed, whereby the instructor regularly sends them a reminder via SMS/text.

Recommendations for Further Research

One area for additional research relates to retention – a topic outside the scope of the present study. Is the vocabulary gained by use of either method here likely to be retained several months later? Do both methods result in similar or different retention rates? The students in both groups in fact took a final vocabulary test two weeks after the post-test which contained items from the Master Vocabulary List, as part of their final course exams, and performed similarly. Most of the studies reviewed took place in a shorter time-frame than the present one, and only one (Hirschel & Fritz, 2013) addressed the issue of long-term retention.

As mentioned, many previous studies that quantified the effectiveness of a mobile app for vocabulary learning failed to account for a number of factors, thus devaluing the results. Despite strict measures taken to account for and control such variables, the present study similarly failed to foresee students in group A choosing another study method over the one given to them, compromising the quantitative portion of the study. It is my strong opinion that the only way an accurate study with an aim of quantifying the relative efficiency of a mobile app versus notebooks, or paper cards, or another mobile app, can actually produce accurate results is if it completely eliminates the outside-class/self-study aspect so

that 100% of study time is in class, and observed. Perhaps such a study needs to be done.

The present study's limitations, failures, and final recommendations invite several other future studies, any of which should be valuable:

1. The algorithms and spacing schemes used by different vocabulary apps.

2. A comparison of another mobile flashcard app versus *Memrise,* focused on performance and preferences.

3. A study that implements the recommendations for *Memrise* use from the present one, to see if the app can truly be used successfully as a vocabulary acquisition tool.

References

4tops, (2011, December 24). Spaced repetition [Blog post]. Retrieved from

    https://artofmemory.com/forums/spaced-repetition-1539.html

Agca, R.K., & Ozdemir, S. (2013). Foreign language vocabulary learning with

    mobile technology. *Procedia and Behavioral Sciences, 83*, 781-785. doi:

    10.1016/j.sbspro.2013.06.147

Aron, A., & Aron, E.N. (2003). *Statistics for psychology, 3rd ed.* Upper Saddle

    River, NJ: Prentice Hall/Pearson Education.

Azabdaftari, B. & Mozaheb, M.A. (2012). Comparing vocabulary learning of EFL

    learners by using two different strategies: Mobile learning vs. flashcards.

    The EUROCALL Review, 20(2), 47-59.

Bahrick, H. (1993). Maintenance of foreign language vocabulary and spacing

    effect. *Association for Psychological Science (aps), 4*(5), 316-321.

Bahrick, H. P., & Hall, L. K. (2005). The importance of retrieval failures to

    long-term retention: A metacognitive explanation of the spacing effect.

    Journal of Memory and Language, 52, 566.

Basoglu, E.B., & Adkemir, O. (2010). A comparison of undergraduate students'

    English vocabulary learning: Using mobile phones and flashcards. *TOJET:*

    The Turkish Online Journal of Educational Technology, 9(3), 1-7.

Pleco Forums, (2017, May 18). Best Setting for flashcard spaced repetition?

  [Msg no. 1-5]. Retrieved from www.plecoforums.com/threads/best-setting

  -for-flashcard-spaced-repetition.5525/

BlueRaja, 2016, August 2. A better spaced-repetition learning algorithm: SM2+

  [Blog post]. Retrieved from www.blueraja.com/blog/477/

  a-better-spaced-repetition-learning-algorithm-sm2

Burgess, S., & Murray, A. (2013). Use of traditional and smartphone app

  flashcards in an introductory psychology class. *Journal of Instructional*

  Pedagogies, 13, 1-7.

Burston, J. (2013). Mobile-assisted language learning: A selected annotated

  bibliography of implementation studies 1994-2012. *Language Learning &*

  Technology, 17(3),157-225.

Choe, A. T. (2016). A critical review of Pimsleur language learning programs.

  Hawaii Pacific University TESOL Working Paper Series 14, 108-120.

  http://www.hpu.edu

Duffy, J. (2017). The best language-learning software of 2017. *PCMag.com*

  https://www.pcmag.com/article2/0,2817,2381904,00.asp

Ebbinghaus, H. (1885). *Über das gedächtnis.Untersuchungen zur experimentellen*

  psychologie. Duncker & Humblot, Germany.

Gu, Y., & Johnson, R. (1996). Vocabulary learning strategies and language learning strategies. *Language Learning, 46*(4), 563-755. doi: 10.1111/j.1467-1770.1996.tb01355.x

Hayati, A., Jalilifar, A., & Mashhadi, A. (2013). Using Short Message Service (SMS) to teach English idioms to EFL students, *British Journal of Educational Technology, 44*(1), 66-81.

Healey, D., & Johnson, N. (1997). A place to start in selecting software. *CAELL Journal 8:1, Winter 1997/98.* Retrieved from http://www.deborahhealey.com/cj_software_selection.html

Hirschel, R., & Fritz, E. (2013). Learning Vocabulary: CALL program versus vocabulary notebook. *Elsevier, Ltd. System,41*,639-653.

Hsiu-Ting, H., & Yu-Chuan J.C. (2012). Vocabulary learning with mobile technology: What students learn and how they react. *World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2012*, 1056–1061.

Hu, Z. (2011). Emerging vocabulary learning: From a perspective of activities facilitated by mobile devices. *World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2011*, 1334–1340.

Hubbard, P. (last updated 2015). An invitation to CALL, Unit 2:

      Finding and evaluating CALL resources. Retrieved from

      https://web.stanford.edu/~efs/callcourse/

Jones, R. (2010). From memory palaces to spacing algorithms: Approaches to

      second-language vocabulary learning. *Language Learning & Technology,*

      14(2), 4-11.

Jung, U. (2015, September 25). Discontent over university evaluations and reform.

      University World News, Issue 00484. Retrieved from

      http://www.universityworldnews.com.

Kent, D., & Sherman, B. (2014). Exploring Memrise application use by Korean

      junior college students studying EFL vocabulary in a blended learning

      context. *STEM Journal, 14*(3), 169-192.

Kornell, N. (2009). Optimising learning using flashcards: Spacing is more

      effective than cramming. *Applied Cognitive Psychology, 23*(9)*,* 1297-1317.

      doi: 10.1002/acp.1537

Lawson, M., & Hogben, D. (1996). The vocabulary-learning strategies of

      foreign-language students. *Language Learning,* 46(1), 1-188. doi:

      10.1111/j.1467-1770.1996.tb00642.x

Leitner, S. (1972). *So lernt men lernen.* Herder, Freiburg, Germany.

Lu, M. (2008). Effectiveness of vocabulary learning via mobile phone. *Journal of Computer Assisted Learning, 24*, 515-525.

Mahdi, S. (2017). Effectiveness of mobile devices on vocabulary learning: A meta-analysis. *Journal of Educational Computing Research*, 55. doi: 10.1177/0735633117698826

Memrise Community Hub (2016). Frequency that words are reviewed? Retrieved from https://community.memrise.com/t/ frequency-that-words-are-reviewed/4454.

McDonough, S. (2001). Promoting self-regulation in foreign language learners. *The Clearing House,* 7(6)*, 323-326. doi: 10.1080/00098650109599218

Miles, S., & Kwon, Chung-Ja (2008). Benefits of using CALL vocabulary programs to provide systematic word recycling. *English Teaching,* 63(1), 199-216.

Mondria, J., & 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.

Mizumoto, A. (2012). Exploring the effects of self-efficacy on vocabulary learning strategies. *Studies in Self-Access Learning Journal, 3*(4), 423-437.

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.

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. doi:

    10.1080/09588221.2010.520675

Nation, P., & Waring, R. (1997). In Schmitt, N. and M. McCarthy (Eds.).

    Vocabulary: Description, Acquisition, & Pedagogy, pp. 6-19. Cambridge

    University Press, U.K. Retrieved from:

    https://www.lextutor.ca/research/nation_waring_97.html

Nation, I.S.P. (2001). *Learning Vocabulary in Another Language*, Cambridge

    University Press, U.K. doi: 10.1017/CBO9781139524759

Nikoopour, J., & Kazemi, A. (2014). Vocabulary learning through digitized &

    non-digitized flashcards delivery. *Procedia - Social and Behavioral*

    Sciences 98, 1366-1373.

Nisbet, D., & Austin, D. (2013). Enhancing ESL vocabulary development through

    the use of mobile technology. *Journal of Adult Education, 42*(1), 1-7.

Oxford, R., & Crookall, D. (1990). Vocabulary learning: A critical analysis of

techniques. *TESL Canada Journal, 7*(2), 9-30. doi: 10.18806/tesl.v7i2.566

Oxford, R., & Scarcella, R.C. (1994). Second language vocabulary learning

among adults: State of the art in vocabulary instruction. *System, 22*(2),

231-243.

Phillips, J. (2009). StudyBlue for iPhone. *MacWorld*. Retrieved from

https://www.macworld.com/article/1142854/studyblue.html.

Pimsleur, P. (1967), A memory schedule. *The Modern Language Journal,*

51, 73–75. doi: 10.1111/j.1540-4781.1967.tb06700.x

Shulman, R. (2017). A polyglot launched a product that is projected to be

Kickstarter's most successful app campaign in history. *Forbes Magazine.*

Retrieved from https://www.forbes.com/sites/robynshulman/

2017/09/22/this-learning-system-is-projected-to-be-kickstarters-

most-successful-app-campaign-in-history/#136405627eec

Son, J.B., & O'Neill S. (2005). Exploring and evaluating language learning web

sites. In *Enhancing learning and teaching: pedagogy, technology and*

language (pp. 215-227). Post Pressed, Flaxton, Australia.

Suwantarathip, O., & Orawiwatnakul, W. (2015). Using mobile-assisted exercises

to support students' vocabulary skill development. *TOJET: The Turkish*

Online Journal of Educational Technology, 14(1), 164-171.

Thalheimer, W. (2006, February). Spacing Learning Events Over Time: What the

> Research Says. Retrieved from http://www.work-learning.com/catalog/

Thornton, P., & Houser, C. (2001). Learning on the move: Vocabulary study via

> email and mobile phone SMS. *World Conference on Educational Media*

> *and Technology 2001*, 1896–1897.

Wu, Q. (2015). Designing a smart phone to teach English (L2) vocabulary.

> *Computers & Education 85*, 170-179.

Wyner, G. (2017). Fluent Forever, the app: Learn to *think in any language.

(2017). Retrieved from https://www.kickstarter.com/projects/gabrielwyner/

> fluent-forever-the-app-learn-to-think-in-any-langu

Appendix A

	Web-based (publicly)	iOS (iTunes) App Store	Android (Google Play)	Algorithm / Spaced Review Method	Other Notes
Anki	Oct. 2006	Unknown/$24.99	2010 / free	Mod. SM-2	One of oldest, most popular mobile apps
Brainscape	2009	2010 / free	2014 / free	CBR – Confidence-based Repetition	Users self-rate "know" scale 1-5
Chegg (Flashcards+)	unknown	2016/ / free	No	unknown	*ChegStudy* app on Android
Cram	2001*	2009 / free	2013 / free	Mod. Leitner system	*as "Flashcard Exchange"
Course Hero	2008	2016 / free	2016 / free	unknown	Uses FlashTutor on Android, Flashcards – Free Apple
DuoLingo (Tinycards)	2012	2016 / free	2017 / free	Mod. SM-2	
iKnow!	2008?	2016 / free	2016 / free	unknown	Jap + Chin lang. Software Japanese co.
MemoryLifter	2001	No app	No app	Mod. Leitner system	Freeware
Mnemosyne	2003	2010? / free	2010? / free	Mod. SM-2	Non-commercial data collect project
OpenCards	2008-Open Office, 2011 –st alone	No app	No app	Mod. SM-2, allows 'priority' label	"Optimal Learn" sched. Uses .ppt slides as cards Open source
Quizlet	2007	2014	2013	"Long-Term Learning"	"Long-Term Learning" for Quizlet Plus only (paid)
StudyBlue	2009	2009	2011	unknown	
SuperMemo	1987-pc soft 2001-online	2010 / free + paid courses	2012 / free + paid courses	SM-17	Lang. software not just a flashcard system
Tiny Classroom	No	No	2010 / free	Unknown	My favorite
Vocabulary Trainer	1997-pc 1999 ?	No	2013? / free + paid courses	5-box Leitner system	aka *VTrain* ; German co.
Memrise	2010	2013 / free	2013 / free	Mod. SM-2 "Plant" + "Water" sessions	Multiple choice defs + fill in multiple levels + pts

Info for the above table was gathered from a variety of online sources, including company websites and blogs, and technical journals/web pages:

https://apps.ankiweb.net/; https://www.ankiapp.com/

https://www.brainscape.com/

http://www.chegg.com/mobile/

http://www.cram.com/

https://www.coursehero.com/flashcards/

https://tinycards.duolingo.com/

https://www.memrise.com/

http://www.memorylifter.com/

https://mnemosyne-proj.org/

http://opencards.info/

https://quizlet.com/en-gb

https://www.studyblue.com/online-flashcards

https://www.supermemo.com/en/frontpage

http://www.tinyclassroom.com/

https://vocabulary-trainer-flashcards.soft112.com/

https://www.vtrain.net/futu0.htm

https://www.facebook.com/pg/Brainscape/about/?ref=page_internal

https://appadvice.com

https://www.maketecheasier.com

http://www.makeuseof.com/tag/6-flash-card-apps-for-android-compared-which-is-the-best

SM-2 Algorithm

Taken from the *SuperMemo* website (https://www.supermemo.com/english/ol/sm2.htm):

SM-2 assigns a decreasing interval time for review based on what type of response it receives on a 0 to 5 scale.

1. With all items associate an E-Factor equal to 2.5.

2. Repeat items using the following intervals:

I(1):=1

I(2):=6

for n>2: I(n):=I(n-1)*EF

where:

I(n) - inter-repetition interval after the n-th repetition (in days),

EF - E-Factor of a given item

If interval is a fraction, round it up

After each repetition assess the quality of repetition response in 0-5 grade scale:

      5 - perfect response

      4 - correct response after a hesitation

      3 - correct response recalled with serious difficulty

      2 - incorrect response; where the correct one seemed easy to recall

      1 - incorrect response; the correct one remembered

      0 - complete blackout.

After each repetition modify the E-Factor of the recently repeated item according to the formula:

      EF':=EF+(0.1-(5-q)*(0.08+(5-q)*0.02))

EF' - new value of the E-Factor,

EF - old value of the E-Factor,

q - quality of the response in the 0-5 grade scale.

If EF is less than 1.3 then let EF be 1.3.

If the quality response was lower than 3 then start repetitions for the item from the beginning without changing the E-Factor (i.e. use intervals I(1), I(2) etc. as if the item was memorized anew).

After each repetition session of a given day repeat again all items that scored below four in the quality assessment. Continue the repetitions until all of these items score at least four.

Appendix B

Basic Info

Name	Memrise (mobile app)	Cost	Free
URL(s)	https://www.memrise.com/	Available download	Google Play , App Store
Developer / Distributor	Greg Detre, Ed Cooke, Ben Whately (Memrise)	Misc. / Req.	Android 4.1 + iOS 10.0 or later

Learning & Teaching

	Comments
Learner level	Appropriate for all levels of English learners
Learner Style / Traits	1. A disciplined learner who will use the app fairly regularly 2. An avg. leaner who does not mind repetition, review
Approach / Method	Multiple choice (choose correct definition) & type word (for definition) using spaced-repetition; an initial 'planting' / learn phase: words are shown then quizzed, followed by a 'watering'/review phase: 'difficult' words (as defined by the spaced-repetition algorithm) are reviewed. The app adds words to 'long-term memory' status after they have been successfully completed and reviewed X number of times.

EVALUATION   1 = Poor   2 = OK   3 = Good

Design/ Technical

Features	1	2	3
Navigation			X
Speed			X
Visual appeal	X	X	
Maintained / Up-to-date			X

	1	2	3
Target culture	X		
Authentic / real-world	X		
Motivating / interesting		X	

Content / Activities

Criteria	1	2	3
Learner-centered			X
Learner goals		X	
Instructions	X		
Levels + sequence		X	

Feedback

OVERALL RATING = 2

is . . . / includes . . .	1	2	3
timely + effective			X
interaction with other learners		X	

82

Good points:

1. Timer makes user choose definition; length of time seems appropriate.

2. Feedback is immediate and easy to follow: an incorrect answer turns red with an X, a correct one turns green.

3. Uploaded images appear automatically in the learn phase, along with the definition

4. User can set a daily goal of 5, 15, or 45 min.

5. All users in a given course can add a 'mem' or memory hint for any word, which can then be seen by other users.   (This would allow for word association/mnemonic devices, etc.)

Bad points:

The point system + levels = make no sense to me! The reason for them is presumably to motivate users to compete against one another. But, how fast you reply to any given question, plus how many times you have to review a word all factor into total points, so that point totals vary widely between users who have completed the same goals as far as number of words/ % of course completed. Plus, I'm not really sure what student would get excited over this point system. Nothing cool happens / there aren't any rewards, even if you are at the head of one of the "leader boards" for the week, month, or all-time. Most competition would still rely on external competition, led by the instructor, for example.

🕐 The color scheme, visual layout, design = terrible! Nothing that would motivate me. Nothing grabs your attention or seems 'cool.' (Students thought it was fine . . .)

🕐 The 'most difficult' words feature – which allows users to see what words they are having trouble with and review them in a different manner and would be one of the strongest points, is only accessible by upgrade to Memrise Pro, which costs $9/mo, $19/3 mo, or $129/lifetime.

🕐 The 'plant'/learn phase for me takes way too long. Even words you know are repeated over-and-over needlessly. I realize this is a highly learner-specific criticism, and only a few learners in the study felt the same way; most thought the repetition was about right. Perhaps the whole scheme/basis of their approach is flawed? With paper flashcards, for example, learners do an initial review of a stack of words (10-20), then proceed to sort out difficult ones. This system takes an entirely different approach, as the initial learn phase includes being exposed to only a few words at a time and immediately being quizzed on them as well – multiple times. For an experienced vocabulary and flashcard user like me, it does not seem natural or ideal. The approach seems to be based on the assumption that most people need to be exposed to a word a dozen times in order for it to make its way into memory, and there is no way a user can adjust the algorithm it follows to allow you to complete a course faster.

Here is the posted review schedule from the *Memrise* user blog site:

1) For any word that you always get correct:

    Plant/learn: review again in 4 hours

    First review: water/review again in 24 hours

    After that, 6, 12, 24, 48, 96, and finally 180 days

1) For any word a learner gets wrong during the planting session:

☝ it will be reviewed again in 4 hours, then 12 hours
☝ After that, it follows the normal schedule if you always get it right.
☝ In some courses, if you plant, but don't water for 24 hours, it adds an extra watering session which follows a 4, 12, 24 hour schedule

1) If you've watered an item a few times, and get it mostly right (a typo), it will come up for review   in 4 hours, then return to the normal schedule.

1) If you get it completely wrong during a watering session, the schedule starts over as if you're planting.

Suggestions for Classroom Use:

☝ Introduce ss to the concept of the spacing effect and spaced-repetition before beginning to use it to memorize vocabulary. Weekly goals for use as a class should be set and checked rigorously. Ss should be shown the 'daily goals' option and encouraged to check one. This way, they may actually use it more regularly as it was designed to be used, rather than all in one-go at the beginning while its still a novelty, or at the end, in order to try to cram for an exam . . .

☝ Make it one option among 2-3 mobile flashcard like apps for outside-class vocabulary learning

Appendix C

Master Vocabulary List |

adjust
aerate
aged
agitate
American breakfast
anchovies
appetite
appetizing
apricot

baste
batter
biscuit (Am + Br.)
blanch
blood sausage
bouillon cube
breading
brewed
broth
bubbles
burgundy (wine)

Cajun (cooking)
cantaloupe (melon)
caramelize
casserole
caterer
charbroiled
charcoal
chestnut
Chicago deep dish (pizza)
Chicken Cordon Bleu
chicken fried steak
chili (American food)
chocolate malt
chocolate no-bake cookies
ciabatta (roll)
cinnamon
citrus
clam
coating
cobbler
coffee (French) press
compartment steam
complement
condiment
consommé

Continental breakfast
convection oven
corkscrew
crock pot
crouton
Cuban (sandwich)
cure (meats)

dash (amount)
date (the fruit)
deconstructed
dehydrated
delectable
delicate
demonstrate
designated
dip
distilled
dragon fruit
dressed (fish)
dunk
Dutch oven

edible
eggplant
El cartoccio
enchilada
English breakfast
Espagnole sauce
even (evenly cooked,
evenly spread)

fare (referring to food)
fattening
faux caviar
ferment
fiber
fillet
fillet mignon
flambé
flapjacks
flash-frozen
foam (for desserts)
freeze-dried
fritter
frosting

generous
ginger
ginseng
gluten
gnocchi
goblet
graham cracker crust
gravy
green onion / shallot
greens (in a salad)
Grenache (wine)
griddle
guacamole (dip)

hand tossed
hands-on
hash
hearty
hoagie roll
horseradish
hot sauce (American)

ice cream sandwich
icing
impression
inflorescent vegetable
infuse
install
inventory

keep an eye on
key lime pie
knead
ladle
Lamington
leavened
legume
lemon grass
lumpia / spring roll
macaroni salad
maintain
mangosteen
marinade
marinate
marmalade
meat jerky
Merlot

mortar and pestle
Naan bread
nutmeg
octopus
open-faced
oven rack
over-easy
oyster

papaya
parbake
parboil
pare
parfait
patisserie
patty
Pepper jack cheese
perishable
phyllo pastry
pico de gallo
platter
plum
portion
powdered
prawn
proof (alcohol)
proof (dough)
Provolone
pudding (Am. + Br.)
pulp
puree

quench
ravioli
reduce (a sauce)
refreshing
refried beans
remove
roe

rolling pin
root beer float
rotisserie
roux

s'more (American food)
salsa
salt shaker
sanitize
savory
scald
scoop out
scrub
sear
seep
sesame
shuck
sieve
sift
silk worm larvae
simmer
sinful
skewer
skim milk
slurp
smidgen
soggy
sourdough
sous-vide
spoil
Sriracha chili sauce
stale
staple (to describe a food/diet)
steep
strips (to count food)
strudel
stuffing
submerge

suds
sunny side up
Swiss (Elemental) cheese

t-bone
Tabasco sauce
take out (verb)
tamarind
taro
tartar sauce
teriyaki sauce
tofu
tongs
tortilla
tuber
turmeric
turn down
turn off
turn on
turn up
TV dinner

veloute
vinegar
walnut
whip up
white zinfandel
wok
Worcestershire sauce

zest

233 Total

Appendix D

Group			A Flashcards Scores & Study Time				
Student	Test 1	Test 2	Increase	Initial Known Words	Final Known Words	Midterm Test A - E 25 pts	Total Est. Study Time Home
17A	17	47	+30 / 176%	59	193	24	5 hrs
16A	12	31	+19 / 158%	79	40	10	3 hrs
7A	21	52	+31 / 148%	111	167	24	4 hrs
11A	14	34	+20 / 143%	41	122	13	10 hrs
15A	23	52	+29 / 126%	99	216	23	5 hrs
14A	22	49	+27 / 123%	76	213	23	24 hrs
1A	25	55	+30 / 120%	113	216	24	2 hrs
19A	23	48	+25 / 109%	71	223	23	24 hrs
10A	23	48	+25 / 109%	74	178	22	4 hrs
9A	19	37	+18 / 95%	41	35	10	3 hrs
6A	28	54	+26 / 93%	90	219	25	2 hrs
3A	24	46	+22 / 92%	114	120	24	2 hrs
12A	23	43	+20 / 87%	97	181	23	3-4 hrs
18A	28	46	+18 / 64%	113	70	24	3-4 hrs
4A	31	43	+18 / 58%	94	198	23	1hr
2A	36	53	+17 / 47%	140	212	23	1.5 hrs
8A	27	39	+12 / 44%	130	200	14	45 min
13A	28	49	+21 / 34%	116	210	24	3 hrs
5A	37	47	+10 / 27%	111	165	23	1hr
Average	24.26	45.95	+22 / 97%	93	167	20.83	5.38

Group			B Memrise Scores & Study Time					
Student	Test 1	Test 2	Increase		Initial Known Words	Final Known Words	Midterm Test A - E 25 pts	Total Est. Study Time Home	
19B	16	41	+25	/ 156%	72	170	19	**9**	**hrs**
9B	17	42	+25	/ 147%	78	120	21		
14B	20	40	+20	/ 100%	70	90	19	**4**	**hrs**
18B	26	49	+23	/ 88%	68	100	17	**5**	**hrs**
4B	27	48	+21	/ 78%	64	100	25	**5**	**hrs**
12B	23	41	+18	/ 78%	36	100	17	**4**	**hrs**
3B	20	35	+15	/ 75%	10	40	17	**3**	**hrs**
7B	27	45	+18	/ 67%	87	150	21	**5**	**hrs**
15B	31	51	+20	/ 65%	83	200	21	**3-4**	**hrs**
13B	20	32	+12	/ 60%	48	50	17	**3-4**	**hrs**
6B	29	45	+16	/ 55%	143	50	20	**5**	**hrs**
16B	24	37	+13	/ 54%	73	50	16	**3**	**hrs**
1B	32	49	+17	/ 53%	140	150	23	**3**	**hrs**
8B	39	55	+16	/ 41%	77	180	25	**6**	**hrs**
5B	27*	37	+10	/ 37%	68	60	12	**7**	**hrs**
2B	39	53	+14	/ 36%	145	190	25	**8**	**hrs**
10B	28	34	+6	/ 21%	109	213	21	**3**	**hrs**
17B	23	27	+4	/ 17%	55	60	15	**4**	**hrs**
11B	28	25	-3	/ -10%	102	60	11	**3**	**hrs**
Average	**26**	**41.37**	**+15 / 64%**		**80**	**112**	**19**	4.67	

88

Appendix E

Group A: Flashcards Questionnaires & Responses

Y/N = Initial questionnaire only

I have used flashcards before.   **Y** (8)   **N** (11)

I have used an English vocabulary app or game on my phone.   **Y** (17) **N** (2)

In the future, I will work in a kitchen with foreign chefs.   **Y** (19) **N** (0)

How Often?
1 =   Never   (결코)   2 =   rarely 드물게   3 = sometimes   (가끔)   4 =   fairly often (꽤 자주)
5 =   often (자주)

	1	2	3	4	5
Initial only: Before this, I studied Eng vocab by myself outside of class.		7	10	2	
Initial: When I see an Eng word I don't know, I find its meaning in the dic			2	9	8
Final: same Q / Kor: 모르는 영어 단어를 볼 때, 나는 그 의미를 사전에서 찾는다				7	12
Initial: I use an app or game on my phone or laptop to study English vocabulary.	2	10	6		1
Final: same Q (휴대 전화 나 노트북에서 앱이나 게임을 사용하여 영어단어를 공부하다)	5	9	3	1	1

To what degree/how much?   어느 정도까지 / 얼마나?

1 = No! 아냐!   2 = Not really/ 별로 안그럴것같다   3 = maybe / 아마 그래 /50-50
4 = Yes, I think so 예, 그런다고   5 =   Yes! 확실히 사실야! /100% agree

	1	2	3	4	5
Initial only: Flashcards seem like a good way to learn vocabulary quickly.		3	8	8	
Initial only: I need to know lots of Eng cooking vocabulary for my future.			2	4	13
Initial only: I will go on an internship next year and work in a foreign restaurant.			3	4	12
Final: same Q (플래시카드는단어를 빨리 배우는 좋은 방법 일것같다)		1	8	7	3
Initial: Flashcards are easy to make and keep with me.	1	3	5	8	2
Final: same Q (쉽게 만들고 들고 다니기 쉽다)		4	6	3	6
Initial: With the Yes/No 2-stack method, flashcards are easy to use.		1	10	6	2
Final: same Q (예 / 아니오 2-더미방법으로서 그들은 사용하기 쉽다)	1	2	4	10	3
Initial: Flashcards are more effective than memorizing words from a list.		2	13	1	3

		1	2	3	4
Final: same Q (플래시카드는 단어 목록에 어휘를 외우는 것보다 효과적이다)		4	5	6	4
Initial: Flashcards are fun to use.		2	9	7	1
Final: same Q (사용하기 재미 있다)			9	5	5
Final only: Compared to a vocab mobile app, flashcards seem a better method. 어휘 모바일 앱에 비해 플래시 카드가 더 좋은 방법 인 것같다.		2	7	7	3
Final Only: I think studying vocabulary with a mobile app would be fun. 나는 모바일 앱으로 어휘를 공부하는 것이 재미있을 것이라고 생각한다		8	6	4	1

Final student comments:
Cards are easy to lose (3) / Better if they had a ring (1)
Easy to memorize vocabulary with them (8) / It takes more time to memorize words (1)
Makes it easy to focus on non-memorized words (3)   Are fun (2) Not easy to use (2) / Don't like to carry them around (3) Making them is convenient (2) / Making them is boring (1)
Writing on them is useful (1)   / They aren't as good as writing in context/a notebook (3)

Group B: *Memrise* **Questionnaires & Responses**

Y/N = Initial questionnaire only

I have used Memrise before.   **Y** (1)   **N** (16)

I have used an English vocabulary app or game on my phone.   **Y** (9)   **N** (8)

In the future, I will work in a kitchen with foreign chefs.   **Y** (15) **N** (1) – 1 N/A

How Often?
1 =   Never   (결코)   2 =   rarely 드물게   3 = sometimes   (가끔)   4 =   fairly often (꽤 자주)
5 =   often (자주)

	1	2	3	4	5
Initial only: Before this, I studied Eng vocab by myself outside of class.		5	6	1	5
Initial: When I see an Eng word I don't know, I find its meaning in the dic	1		3	8	5
Final: same Q / Kor: 모르는 영어 단어를 볼 때, 나는 그 의미를 사전에서 찾는다	1	1	3	8	3
Initial: I use an app or game on my phone or laptop to study English vocabulary.	5	3	5	2	2
Final: same Q (휴대 전화 나 노트북에서 앱이나 게임을 사용하여	2	6	4	4	

		영어단어를 공부하다)					

To what degree/how much? 어느 정도까지 / 얼마나?

1 = No! 아냐!   2 = Not really/ 별로 안그럴것같다   3 = maybe / 아마 그래 /50-50
4 = Yes, I think so 예, 그런다고   5 = Yes! 확실히 사실야! /100% agree

	1	2	3	4	5
Initial only: I need to know lots of Eng cooking vocabulary for my future.			1	8	8
Initial only: I will go on an internship next year and work in a foreign restaurant.		1	2	5	9
Initial: Memrise seems like a good way to learn vocabulary quickly.			4	8	5
Final: same Q (Memrise는 단어를 빨리 배우는 좋은 방법 일것같다)		1	6	6	2
Initial: It was easy to create an account/register, and start using Memrise.			5	7	5
Final: same Q (계정을 만들고/온라인으로 등록하고 사용하기 쉽다)		1	2	8	4
Initial: I like the way Memrise works.		1	7	5	4
Final: same Q (작동하는 방식을 좋아한다)			2	10	3
Initial: Memrise is more effective than memorizing words from a list.		1	2	9	5
Final: same Q (단어 목록에 어휘를 외우는 것보다 효과적이다)			2	11	2
Initial: It's fun to use.			7	6	4
Final: same Q (사용하기 재미 있다)		1	7	4	3
Final only: Compared to paper flashcards, this app seems a better method for study. 종이 플래쉬 카드에 비해,이것은 공부기위해더 좋은 방법 인 것 같다				11	2
Final Only: I think studying vocabulary with paper cards would be fun. 나는 종이 카드로 어휘를 공부하는 것이 재미있을 것이라고 생각한다	3	4	2	4	2

Final student comments: *You can use it to study English anytime, anywhere. (1)*
It's easy to study with it. (2)
It's fun to use. (1)

Appendix F

General Questions

Did you use (flashcards/Memrise)? About how many hours in all?   How often?
(Elicited use pattern (with hands / used Korean language as necessary).
그것을 사용 했어요? 모두 몇 시간 쯤? 얼마나 자주?

Did you think it was a good way to study vocabulary?
어휘를 공부하기위해 좋은 방법이라고 생각 했어요?

What were the bad points?   What are things you did not like about it? 나쁜 점은?

What were the positive points? What are things that you liked? 좋은 점은?

Comparing this method to notebook study or your own method – which do you think
would be best for you? Why? 선택할 수 있다면 / 이 방법을 다른 것하고 비교하면,
어느 것은 당신에게 가장 좋을 것이라고 생각해요? / 딴 방법?

Do you think you will continue to use this method in the future if you need to study
vocabulary for any class? 앞으로, 수업을 위해 어휘를 공부해야하는경우, 이방법을
사용할것이라고 생각해요? Why/Why not?

Responses (by Group)
A-1

Hours: 10, 5, 3, 5

Cards / Other methods?   cards only* (17A = greatest improvement), "writing"
                    (in notebook/English word + def), cards + writing, "writing"

Why writing?   favorite way / ran out of cards (after 40)

Cards good for vocab?   yes, so-so, yes, yes

Bad points?   making cards boring, lost cards, hard to make w/o
                    card paper, takes long time to study w them

Good pts?   Can remember for long time, fast to remember words, handy,
nothing good

Use pattern?   wk 8 little, wk 9 – lots, some wk 1 + wk 9,
1 wk before test, used cards before midterm then notebook before final

Compared other methods?   better, better than writing, same as writing, writing better

Paper or app?   use phone other things so no app, app handier,
app, app

A-2

Hrs?   2, 24, 4, 3, 24   (24 studied together)

Pattern?   1 wk before final, mostly before final test, 10 min before every
class + 2-3 hrs before the final test, 24=1 wk before exam every day

Methods?   Word list, word lists - did not quiz each other, word list, word list

Why word list?   Like writing words down, can see many words at same time with
list, like to see more than 1 word at a time, made 40 cards but
made own list

Done in Korean class?   No, just my style / Not really / used in Korean classes / not
used before

Why cards not useful?   Word list more comfortable, using list is a habit, lost cards

Cards/app/word list best for you?   Word list, list + app, word list, word list, word list

A-3

Hrs?   2, 2, 3-4, 3-4, 3

Cards/other methods?   Cards + note paper (list), writing in notebook, used cards
as shown, wrote on list and said words aloud, word list

Bad pts?   Lost cards (2) , nothing bad, uncomfortable carrying them

Why use word list?   Can remember with list, someone taught/learned it from
teachers

Pattern? Wk 1 study + lot before test (2), all before the test (3)

Cards, word list, app? Writing words in list (2), writing list + app, app – easy use, app + reading list

A-4

Hrs? 30 min, 1.5 hrs, 30min – 1 hr, 2hrs, 1 hr

Pattern? all at end before test, before test, 30 min before test + little before that, all before test, most before test + little before that

Cards/other methods? printed list found meaning + wrote it, both cards + list, used cards but for ones didn't know wrote meaning in notebook, printed list + wrote words out list, read words from list

Why? easy way to learn,   used to doing that way, in Jr. High school was taught, hard to make flashcards, studied that way in past

Bad pts? Lost cards, lost some cards, hard to keep/carry, can't write on them in hand like phone, need ring like ones in (Korean) store

List/cards/app? App, writing in list (3), writing own words in app good but not good if words already chosen, difficult to use app because I'd chat with it (3), think writing is the best way to memorize words

---

Group B: *Memrise* **App**

B-1

Hrs? Less than 10, 4, 5

Pattern? *every week 30 min, 1 hr every 2 wks (highest performance), all in 1st/lastweek

Like app? Good? Yes, easy to memorize, liked repeating nature, fun

Bad pts?   Too many words in course, takes too long to finish, typing words in

Repetition?   a little much, just enough, a little much

Other methods?   Used word list on phone, app only, word list in notebook (wrote),

Use in future?   Yes – for sure

App/notebook/other? App (2), YouTube

B-2

Hrs?   4, 3, 3-4, 3-4

Pattern?   Most before test, steady increase, At beginning + most before test (2), M-shape use

Like? Not like?   Loved it, okay, words in course were good (2), simple but hard to progress to next level, easy because it's on phone

Repetition?   Enough (just right) – everyone

App/other methods?   App only (everyone)

App/list/other?   Memrise app (everyone)

Other comments?   Ads on free version – Hated them, pop-up to upgrade didn't like (only noticed ads a little)

B-3

Hrs?   5, 3, 5, 5

Pattern?   All – week 1, then at end before test

Best thing?   Easier to memorize words with it, easy to bring/take (convenient)

Bad things?   Not fun – because it is still studying

Ads?   Ads no problem (everyone)
Typing in words not a problem, but better if just multiple choice

App/NB/other?   App! - everyone

B-4

Hrs?   8, 6, 7, 3

Pattern?   1 hr every two weeeks, here and there unbalanced, at end before test, Finished all of it first 2 weeks (2)

Good pts?   Hate phone apps, nothing good, app better than other ways of study, better than using notebook, will remember words learned on Memrise longer

Bad pts?   Definitions too short (lack of context), classic review is no good
Layout – gives you 6 choices so even if you know word on test, it's difficult to pick out on the phone

Repetition?   Just right (enough) – everyone

App/NB/other?Learn from reading in context better (2), app, app

B-5

Hrs?   3, 3-4, 4

Good way?   Yes – fun, easy, explanations were easy

Difficult/bad? Makes you spell the word = difficult on phone, nothing bad

Design?   Good, ok, no problem

Pattern?   At beginning, beginning 2 hrs, middle 30 min, 2 hrs at end, almost all the first week, beginning + end

Why?   It wasn't homework so didn't use it more, studied at beg because   it was new, don't want to study outside class

Use in future?   Yes – all

App/List/other best? This app (2), use word list – seems easier than app

Appendix G

Initial Memrise Data

Student	Study Time last 30 days	Words Learned	Long-Term Memory	Difficult Words	% of Course Completed	Points last 30 days
	9 min	213	48	0	100	7836
10B	14 min	105	0	20+	49	4182
9B	1 hr 12 min	86	0	20+	40	37153
6B	0	84	0	13	39	0
12B	1 hr 6 min	77	28	20+	36	19556
14B	17 min	77	6	20+	36	4726
19B	7 min	65	0	7	30	962
17B	17 min	63	0	20+	29	4456
7B	7 min	49	0	5	23	3252
15B	21 min	48	0	15	22	11194
11B	0	26	0	13	12	0
5B	9 min	25	0	9	11	4583
3B	0	0	0	1	0	0

Data for the month of October, collected on November 1st. "Study Time" and "Points" are for the previous 30 days (Oct.); the remaining categories are cumulative, representing the period from the beginning of the present study in August through the end of October.

Final Memrise Data
collected on 16 Nov, for the period Oct 18 – Nov 16. Final achievement.

Student	Study Time last 30 days	Words Learned	Long-Term Memory	Difficult Words	% of Course Completed
8B	30 min	213	118	1	100
18B	*no data*	*213*	*no data*	*no data*	*100*
2B	*no data*	*213*	*no data*	*no data*	*100*
12B	3 hrs 54 min	132	68	20+	61
4B	*no data*	*131*	*no data*	*no data*	*61*
1B	*no data*	*117*	*no data*	*no data*	*55*
15B	1 hr 23 min	117	11	20+	54
10B	22 min	106	0	20+	49
9B	1 hr 42 min	100	0	20+	46
14B	1 hr 10 min	93	77	20+	43
6B	2 min	88	0	13	41
19B	1 hr 21 min	86	53	20+	40
17B	19 min	65	0	20+	30
7B	7 min	49	0	5	23
11B	44 min	42	7	20+	19
5B	46 min	75	0	20+	35
3B	28 min	42	0	10	19
13B	*no data*	*61*	*no data*	*no data*	*29*
16B	*no data*	*57*	*no data*	*no data*	*27*