Research paper

Digital flashcard L2 Vocabulary learning out-performs traditional flashcards at lower proficiency levels: A mixed-methods study of 139 Japanese university students

Robert John Ashcroft*, Robert Cvitkovic* and Max Praver**
*Tokai University, Japan | **Meijo University, Japan
_____
*bob.ashcroft@tsc.u-tokai.ac.jp | **bcvitkovic@yahoo.ca | *** praver@meijo-u.ac.jp

Abstract

This study investigates the effect of using digital flashcards on L2 vocabulary learning compared to using paper flashcards, at different levels of English proficiency. Although flashcards are generally believed to be one of the most efficient vocabulary study techniques available, little empirical data is available in terms of the comparative effectiveness of digital flashcards, and at different levels of student English proficiency. This study used a mixed-methods experimental design. The between-subjects factor was English Proficiency consisting of three groups: basic, intermediate and advanced. All participants underwent both a digital flashcards treatment and paper flashcards treatment using words from the Academic Words List. For each study mode, the two dependent variables were Immediate, and Delayed Relative Vocabulary Gain. The results of this study indicated that Japanese university students of lower levels of English proficiency have significantly higher vocabulary learning gains when using digital flashcards than when using paper flashcards. Students at higher levels of proficiency performed equally well using both study modes. It appears that by compensating for the gap in metacognitive awareness and effective learning strategies between students of lower and higher levels of language proficiency, digital flashcards may provide the additional support lower-level learners need to match their advanced-level peers in terms of their rate of deliberate vocabulary acquisition.

Keywords: Vocabulary, digital flashcards, paired-associates, autonomy, English proficiency, Academic Words List.

1. Introduction

The exponential growth and development of computer technology is having a significant impact on many aspects of foreign language pedagogy. Most teachers intuitively recognize the opportunities afforded by Computer Assisted Language Learning (CALL) materials and strive to integrate these technological innovations into their teaching practices. However, due to the rapid rate of change, related research and accompanying pedagogy can often lag behind the development of new CALL applications. As a result, teachers may lack the support of a theoretical framework. Moreover, digital technology often appears intrinsically desirable in itself, rather than because of the possible objective benefits to students. This superficial appeal, combined with a lack of pedagogy, can result in unrealistic expectations of CALL applications in terms of learning outcomes (Gartner, 2017). It is difficult for language teachers to recognize which CALL applications will enhance student learning and which will not. Although the influence of CALL is felt throughout most aspects of language teaching and learning, the present study specifically examines vocabulary learning.

One way in which CALL technology has influenced the field of L2 vocabulary learning has been with the emergence of digital flashcards, with Quizlet and Anki being popular examples. Despite the widespread use of these applications among students and teachers, the comparative effectiveness of digitized flashcards remains under-

researched (Nation and Webb, 2011). The authors of the present study used the Quizlet application in an effort to determine the effectiveness of digital flashcards for learning L2 vocabulary compared to the traditional paper variety at different student proficiency levels. The results of the experiment are summarized in the five points below.

- Immediate vocabulary gains are significantly higher for digital flashcards than paper flashcards for both intermediate and lower levels of English proficiency.
- There is no significant difference between digital flashcards and paper flashcards for immediate vocabulary gains for higher English proficiency levels.
- Delayed vocabulary gains are significantly higher for digital flashcards than paper flashcards for students of lower levels of English proficiency.
- There is no significant difference between digital flashcards and paper flashcards for delayed vocabulary gains for intermediate English proficiency levels.
- Delayed vocabulary gains are significantly higher for paper flashcards than digital flashcards at higher levels of English proficiency.

2. Literature review

The following sections describe the relative merits of using paper and digital flashcards for vocabulary learning. The description includes a detailed consideration of how digital flashcards might further enhance the benefits inherent in paper flashcards. This is followed by a summary of a number of studies into the relative effectiveness of CALL and traditional vocabulary learning.

2.1. Paper flashcards

Research suggests that using paper flashcards is one of the most efficient means of deliberate vocabulary study techniques available (Elgort, 2010). Also known as paired-associate learning, this technique involves using small cards with the target L2 word on one side and the meaning of that word on the other. Using flashcards is thought to be particularly effective due to a combination of factors. First, because flashcards are portable, and therefore convenient, they can help engender student autonomy (Nation, 1995, 2003, 2005). The freedom to study whenever and wherever they like can have a liberating effect on students. Second, flashcards facilitate spaced-learning (Nation, 2003) where students can revisit items over an extended period (Hulstijn, 2001; Webb, 2007). The positive effect of spaced-learning on vocabulary acquisition is thought to be particularly strong. A further advantage of flashcards is that they can be grouped into sets (Cohen, 1990) based on relevant criteria such as lexical groups or test items. Finally, flashcards can include L1 translations, providing a visual link between L1 and the target language (Cross & James, 2001) and thereby further adding to their positive motivational effect. Although the meaning of the target word can be conveyed in several ways, such as a picture, L2 definition or L2 synonym, research indicates that L1 translation is the most effective method (Laufer & Shmueli, 1997; Nation & Webb, 2011). Using flashcards is more convenient, allows spaced-learning, and can include L1 synonyms of target vocabulary.

There are several points to bear in mind when trying to optimise the effectiveness of vocabulary flashcards. Baddeley (1990) stresses the importance of the retrieval process. Once a new word and its L1 meaning have been met, at the next meeting the student should see only the target word and try to recall the L1 meaning. It is also important to continually change the order of the flashcards (Nation & Webb, 2011). This prevents previous items from triggering the memory of subsequent ones, and also allows students to focus on the more difficult items. The optimal number of items for a study set is also an important consideration. Suppes and Crothers (1967) found that for lower level students it should be around 20 cards, and for more advanced learners, up to 50 cards is acceptable. Using paper flashcards to learn vocabulary is a long-standing technique and has been thoroughly researched. However, a more recent development is the emergence of computer-based, digital flashcard applications.

2.2. Digital flashcards

Several Web 2.0 flashcard applications now allow users to create, study and share with digital flashcards. The digital flashcard application chosen as the focus of this experiment was Quizlet, a popular choice for many students and teachers. The site has more than 20 million monthly users and over 140 million freely available user-made flashcard sets (Quizlet, 2017). The application has an attractive, intuitive interface and requires little set up or computer know-how to start studying new words. The website allows teachers to create a virtual class, and invite students to join. Once students have joined a virtual class they have access to the study sets within the class and can track their progress and that of other members of the class. Teachers have access to information about the study behavior and performance of the class members. There is also a Quizlet app available to download to a mobile device both for Android and i-OS. The app allows flashcard sets to be downloaded to the device and used with or without Internet connection.

The Substitution Augmentation Modification Redefinition (SAMR) Model (Puentedura, 2012) provides a means of assessing the integration of technology and its effect on teaching and learning. It divides CALL innovation into four stages of progressively greater degrees of enhancement. A study by Ashcroft and Imrie (2014) used the SAMR Model to assess the impact of using Quizlet vocabulary flashcards compared to paper vocabulary flashcards. They concluded that digital flashcards might be more effective due to additional features such as audio, immediate feedback, a seamless and user-friendly interface, and their high accessibility through a range of platforms. This additional functionality of Quizlet for L2 vocabulary study, compared to traditional flashcards is discussed in detail below using a framework adapted from Reinders and White (2011) outlining areas in which CALL materials in general can have pedagogical advantages over traditional teaching materials.

2.2.1. New activities

New types of activities are made possible using CALL applications which would be difficult or even impossible with traditional teaching materials. Indeed, this can be said of the Quizlet website, which offers a choice of four study modes. Firstly, the Flashcard Mode includes automated audio rendition of words on the cards. Next, the Learn Mode presents one side of a flashcard and requires the hidden item on the reverse of the card to be entered using the keyboard. If the target word is typed correctly, the program moves on to display the next card. If not, the answer is given, and learners are required to retype the word. The third study mode is Test, where users can set test parameters, such as the number and type (multiple choice, written, true/false, or matching) of questions. The test is generated based on the parameters and users are required to complete the test using the keyboard and mouse. When the test is complete, the total score is displayed, along with a list of the test items including the students' responses, the correct answer, and whether students answered each item correctly or not. The last study mode is called Spell. Here users must enter the target item using the keyboard based on an audio prompt.

In addition to the study modes, there are also two game modes. The first game is receptive. The user must match paired associates against the clock. The app then challenges users to beat their best time. The other game is a productive activity where users must enter the target item when prompted by one half of a paired associate. This must be done before an asteroid crashes into the planet below. Points are awarded for pairs successfully matched. Asteroids fall progressively faster, increasing the difficulty of the activity the further users progress. There is clearly a much greater variety of activities available to vocabulary learners using Quizlet compared to paper flashcards.

2.2.2. Feedback

Immediate feedback dependent on users' input is possible when using CALL materials. The Quizlet app provides high density feedback to users. For example, in the Learn study mode, the app will signal whether an item has been entered correctly or not. If users type in the answer for a different card, there is a confusion alert. A message appears explaining the problem and displaying the card correctly matching the entered response. When learners have worked through all the cards in a study set, all items are

then displayed, indicating whether each one was answered correctly or not. A percentage correct total is also shown. Learners are then required to repeat the process for those items answered incorrectly in the previous round. This process is repeated until the correct total reaches 100 percent. The Quizlet app allows for far richer and more immediate feedback than using paper flashcards.

2.2.3. Non-linearity

Traditional foreign language classrooms typically progress in lockstep (Richards & Schmidt, 2002) fashion with all students transitioning together from one activity to the next under the supervision of the teacher. However, CALL materials offer individual students many more study choices and the freedom to use these in any order and for as long as they wish. This holds true for Quizlet, with four study modes and two game formats. Students can approach learning by using any of the modes and in any order they choose. Moreover, in all modes, Quizlet presents cards to the user in randomized order, an important factor that maximizes vocabulary learning (Nation & Webb, 2011). In addition, flashcard sets can be combined, and individual flashcards can be starred to focus on more challenging words.

2.2.4. Monitoring and recording progress

Many CALL applications have the capability of monitoring and recording progress. This information can be made available to teachers, allowing those students not doing the work to be identified, if necessary. Monitoring data can also be used by the application itself to modify future studying activity. If a student is a member of a Quizlet virtual class, the website tracks the user's performance, and this data is available to students and teachers. The site allows flashcards to be sorted according to users' past performance. The cards are displayed in order from those most to those least often missed. This provides the opportunity for students to reflect on the learning process, and to target more problematic vocabulary. In addition, in the game mode Asteroids, Quizlet recycles words which students have missed earlier in the game.

2.2.5. Control

CALL materials provide a greater degree of control for students than traditional materials. Quizlet allows the option of studying with a subset of cards which have been missed by learners in previous study sessions. This ability to target words based on individualized feedback provides a greater sense of control over the learning process for students. The availability of Quizlet on a personal computer, tablet, or smart phone also passes greater control to users. Increased levels of control provide opportunities for the development of metacognitive skills and learner autonomy (Reinders and White, 2011). The Quizlet application provides additional functionality and control previously unavailable through traditional analogue flashcard use.

2.3 Paper versus digital flashcards: existing research

Although the benefits of the additional functionality of Quizlet over paper flashcards seem apparent, the results from existing empirical studies which examine the relative effectiveness of digital flashcards over their traditional counterparts remain inconclusive. One study of 226 Japanese high school students examined the comparative effect on vocabulary gains of using word lists, word cards and a CALL application to study ten vocabulary items (Nakata, 2008). The experiment found no significant difference between using paper flashcards and the computer application. A further study by Lees (2013) also found no difference between the effectiveness of paper flashcards and Quizlet flashcards. Another study by Hirschel and Fritz (2013) used CALL-based vocabulary learning and vocabulary notebooks with 140 university students in Japan. The results showed no significant difference between the two study modes. A further study also found no significant difference between using paper flashcards and internet based digital cards, however the results did show a significant difference between paper flashcards and digital flashcards available on a mobile device, such as a smart phone or tablet (Nikoopour, Jahanbakhsh & Azin, 2014). In all these studies, participants included only those from within the same English proficiency level band, or mixed level homogenized groups, so none of the results could take into account possible effects of proficiency level on the relative effectiveness of the

treatments. The present study attempts to address this gap in the research by answering the following research question:

RQ1. Does student English proficiency level influence the relative effectiveness of digital and paper flashcards in terms of L2 vocabulary learning gains?

3. Method

The purpose of this study was to investigate any difference in effect of using digital flashcards compared to paper flashcards, and to determine whether students English proficiency level also influenced the effectiveness of either study mode. Participants underwent both digital and paper flashcard treatments. For each treatment, a pre-test was used to determine how many of the target words were already known to each participant, and a post-test indicated how many items had been learned due to the treatment. A delayed post-test measured the rate of attrition of this learning. Details of how the experiment was carried out are provided in the subsections which follow.

3.1. Participants

The participants were 139 native Japanese, English language undergraduate students at a large university in Japan. Ages ranged from 18 to 24 years old, with 64 male and 75 female participants. All participants had received formal English instruction for at least seven years. Students belonged to either basic ($n = $ 32), intermediate ($n = $ 46) or advanced ($n = $ 61) level integrated skills-based English classes. Students had been placed at either basic, intermediate or advanced-level based on a university administered TOEIC listening and reading test, compulsorily taken by all students at the start of their freshman year. A total of seven classes participated in the research: two basic-level classes, two intermediate, and three advanced-level classes. All participants owned a smart phone (either iPhone or Android). Many of the students participating in this research were also enrolled in different English courses through the university during the experimental period. The TOEIC listening and reading score ranges are shown in Table 1.

Table 1. Classes, between subjects groups, and English levels of the participants (N =139)				
Class Level	**Class #**	**Class *n***	**TOEIC (L&R)**	**Level *n***
Basic	1	15	under 230	32
	2	17		
Intermediate	3	21	230 to 550	46
	4	25		
Advanced	5	21	over 550	61
	6	17		
	7	23		

3.2. Design

This study used a mixed-methods experimental design. The within-subjects factor was Study Mode, of which there were two levels, digital flashcards and paper flashcards. The between-subjects factor was English Proficiency, which consisted of three levels, basic, intermediate and advanced. The two dependent variables were Immediate and Delayed Relative Vocabulary Gain. These were defined as the number of new words learned from a closed set of target words expressed as a proportion of those words unknown prior to treatment. Word knowledge was measured using a productive L2 measure which was

prompted with the L1 half of a paired associate, along with the first letter of the L2 target word. A productive measure of vocabulary was chosen because many of the students taking part in the research were also taking English academic writing classes. The authors concluded that a productive measure of vocabulary gains was more appropriately matched to the current academic needs of the students.

3.3. Target vocabulary

A fixed and relatively small set of words was used for several reasons. Firstly, this helped the experiment to reflect the targeted nature of vocabulary study using flashcards in real-world learning contexts, thereby increasing the ecological validity of the design. In addition, measuring learning gains using achievement pre- and post-tests would allow a more precise measure of progress. Measuring changes in overall vocabulary size would, in contrast, be much more problematic since incremental gains would be proportionally very small. A further reason for the use of a small number of specific items was that the treatment period could be kept comparatively short, thereby minimizing the probability of students meeting target words outside the treatments.

The selection of words was informed according to two criteria. Firstly, it was important that items should be largely unknown to the participants so as to allow the effect of the treatments to be measured. Secondly, to maintain ecological validity, it was important that the items were relevant and useful for students to know. Using words from the Academic Words List (AWL) (Coxhead, 2000) allowed both requirements to be accommodated. The AWL was created through the analysis of a corpus of around 3,500,000 running words of written academic text. The list contains 570 word families, divided into nine sub-lists of 60 and a tenth sub-list of 30. Word families are ordered from the most frequent (List 1) to the least frequent (List 10). Coxhead used the most frequent form of each word family, per the academic corpus, when compiling the AWL.

A total of 120 words were used in the present study. For both treatments (digital and paper flashcards), participants studied with a different set of 60 target words. AWL Sub-lists 1 and 2 (representing 120 word families) were selected for this purpose as they are the most frequent AWL words and therefore most likely to be useful for students to know. Using computer randomization software, thirty words were selected at random from List 1, and then combined with 30 randomly selected words from list 2. These 60 words constituted the paper flashcards study set. The remaining 60 words were used for the digital flashcards study set. Randomly selecting words in this way ensured that the two study sets had comparable mean frequencies in the academic corpus, and thereby removed any distorting effect of frequency differentials across study sets.

3.4. Instrumentation

Six dependent measures were administered to each participant. Pre-post, non-identical, 30-item vocabulary tests were developed to measure vocabulary gains for each treatment. For the pre-tests, test items were created for 30 words selected at random from each study set (60 total words). The remaining 30 words in each study set were used for the-post test items. Pre- and post-test scores (each out of 30) were used to calculate the relative vocabulary gains from the study set as a whole. The delayed-post tests were made by selecting 15 items at random from the corresponding pre- and post- tests.

The prompts on the pre- and post-tests included the Japanese equivalent of the target item, as well as a sentence in English with the target word omitted (see Table 2). The first letter of the target word was provided to discourage participants from producing synonyms for target answers (Hughes, 2013). The tests were administered in paper format to allow flexibility for misspelt answers, and for British / American English variations. As a rule, two letters per item were permitted to be misspelt. Both American and British spellings of answers were accepted. The L2 words from the flashcards of the respective treatment were the only acceptable answers.

Table 2. Sample test item

L1 prompt	prompt	answer
財政	He studied banking and (f_____) at business school.	finance

Using different items for pre- and post-tests ensured that learning of target vocabulary due to taking the pre-test and the consequent distorting effect on the results of the post-test could be avoided. However, because the items on the pre-test differed from those on the corresponding post-test, and delayed post-test, it was necessary to check that the tests were of equivalent difficulty. In order to do this, the two sets of pre-, post-, and delayed post-tests were administered to 176 native Japanese students of English studying at the same university as those in the main experiment. The test validation group received no treatment. Like the experimental procedure, the digital cards pre-test was administered in class 1, and the post-test in class 3. The paper cards pre-test was done in class 4 and the post-test in class 6. All participants therefore took all six tests. During the intervening class time, students did not study any vocabulary from any of the tests. Paired-samples *t*-tests were conducted between each of the six combinations of tests. The *t*-tests included a Bonferroni correction to offset for the number of tests. No significant difference in the mean scores for any combination of the tests was found (see Table 3). This result indicated that the tests were of equal difficulty and that any difference in the test scores in the main experiment should be attributed to the intervening treatment.

Table 3: Pre-, Post- and Delayed-Post Tests Validation (n=176)

Combination	t (350)	p	M	SD
QPre/QPost	-0. 543 a	0.59*	-0.07	1.81
QPre/PPre	0. 000 b	1.00*	0.00	1.62
QPre/PPost	.208 b	0.84*	0.03	1.81
QPost/PPre	.334 b	0.74*	0.07	2.94
QPost/PPost	.429 b	0.67*	0.10	3.17
PPre/PPost	.131 b	0.89*	0.03	2.88

a: df=75; b: df=174; *:p>.05

3.5. Treatments

The research was conducted over the course of two semesters. The experimental design included two treatment cycles: a paper flashcard treatment and a digital flashcard treatment. Each treatment cycle spanned three 90-minute classes, making a total of six classes for both treatments combined (see Table 4). This was a repeated measures design, with all participants receiving both treatments. However, the order of treatments was varied. Participants were arbitrarily divided into two groups: The Digital First group (*n*=75: 17 basic, 25 intermediate, and 23 advanced) received the digital flashcard treatment first, followed by paper flashcards. The Paper First" group (*n*=74: 15 basic, 21 intermediate, and 38 advanced) worked with paper flashcards first,

followed by digital flashcards. Adopting this counter-balanced design helped to minimize the effect of order of study mode on the results.

Table 4. Repeated measures experimental design protocol

Class #	Digital First Group	Paper First Group
1	Digital Flashcards Pre-test Quizlet Orientation	Paper Flashcards Pre-test; Writing out Flashcards
2	Digital Treatment Homework assigned	Paper Treatment; Homework assigned
3	Digital Flashcards Post-test	Paper Flashcards Post-test
4	Paper Flashcards Pre-test Writing out Flashcards	Digital Flashcards Pre-test Quizlet Orientation
5	Paper Treatment Homework assigned	Digital Treatment Homework assigned
6	Paper Flashcards Post-test Digital Flashcards Delayed Post-test	Digital Flashcards Post-test Paper Flashcards Delayed Post-test
9	Paper Flashcards Delayed Post-test	Digital Flashcards Delayed Post-test

Each flashcard, both paper and digital, had a target English item on one side and the corresponding Japanese (L1) translation of the most frequent academic meaning according to the Oxford English dictionary (Soanes, 2010) on the other. The English was translated into Japanese by a native Japanese university English teacher. The translations were then confirmed by a second Japanese native of high English proficiency. The prompts on the pre- and post-tests included the same Japanese equivalent of the target item (see Table 2). For both treatments, there was a pre-test, a post-test after the treatment, and a delayed post-test 3 weeks after the treatment. The post-test was included in an effort to measure comparative differences in the rates of attrition of vocabulary learning with the two learning modes, and at different levels of proficiency. A three-week gap between immediate post and delayed post-tests was chosen to reflect what might be a realistic period for the participants between deliberate vocabulary study and the opportunity to use these words again incidentally in the course of their studies.

3.5.1. Digital treatment

The digital flashcard treatment sessions were conducted in a CALL classroom. Students used the Quizlet website, not the downloadable app version. Each student had the use of a computer with Internet connection and a pair of headphones. Additionally, each pair of students could see a monitor showing the teacher's computer screen display. The digital flashcards were prepared on the Quizlet site in advance of the treatment. The cards were organized into three sets of 20 items. In the first 90-minute class of the treatment, the instructor introduced students to the Quizlet application. Students registered with the Quizlet website, and then joined a virtual class (see section *Digital Flashcards*), created by the researchers in advance. Using a prepared set of 20 flashcards based on items taken from the AWL List 3 (i.e., not the items being used for this research), students practiced using the site. These flashcards were prepared with the English item on one side, and the Japanese equivalent on the other. The class was conducted in lockstep (Richards & Schmidt, 2002). Students first examined the flashcards and then were shown how to operate each of the various study modes using the shared monitors. The teacher also demonstrated the audio function available in some of the study modes, which the students then went on to use. During this stage,

the teacher periodically displayed summaries of class progress on the shared monitors. For the last 20 minutes of class, students were shown how to download the Quizlet app to their smart phone. There was no homework assigned, and the deck of 20 items was removed from the virtual class after this session.

The next class began with the pre-test for the digital flashcard word set. Then the students studied with digital flashcards for the remaining 90 minutes using the three sets of 20 items from the digital flashcard study set. Students were free to use any study mode and sets of digital cards in any order they wished. The teacher monitored participants closely, offering assistance and ensuring they remained on task. At the end of the session, students were told how to download the digital flashcard study sets to their mobile phones. For homework, students were told to study the digital flashcards study set with Quizlet using a PC or on their smart phones in preparation for a test to be given next class. No attempt was made to measure Quizlet usage outside class. In the next session, students took the digital flashcard post-test. The digital flashcards delayed post-test was given exactly three weeks after the digital flashcards immediate post-test was administered.

3.5.2. Paper treatment

All paper flashcard treatment sessions were conducted in a standard classroom with whiteboard and moveable chairs and desks. Students were not permitted to use their cell phones. Firstly, the pre-test for the paper flashcards words set was administered. Students were then each given a set of 100 blank paper flashcards measuring 5cm by 2.5cm, bound by a plastic ring. The ring could be detached enabling the cards to be separated. The students were given a copy of the paper flashcard study set. This was on a sheet of A4 paper depicting a table with the 60 target vocabulary items in English in the first column and the corresponding Japanese equivalents in the adjacent column. Students copied the vocabulary onto their set of flashcards. Each card had the English word on one side and the Japanese equivalent on the other. Students then wrote their names on the top cover card of their set and the teacher collected all the flashcard sets and vocabulary tables.

In the following session, students were handed back their paper vocabulary card sets which they had made during the previous class. Students removed the plastic clip from their set of cards, and worked individually to separate the items into two sub-sets: the words they felt they understood, and the words they did not. This task allowed them to focus their efforts on those words which were unfamiliar to them. The students were given 20 minutes to memorize these words. For the next 20 minutes, students worked with a partner to test each other on the words. Pairs took turns to prompt each other with the English (L2) items to elicit the Japanese (L1) meaning. Having the students retrieve the target L1 item from memory using the L2 paired associate is thought to optimize vocabulary learning (Baddeley, 1990). Finally, students were arranged into small groups of four or five members. Students took turns prompting the other students in the group with L2. The first group member to give the correct L1 meaning for each item was awarded the corresponding card. The winner of each round was the student who had received the most cards. This stage lasted for 40 minutes. At the end of class, students were told to use their set of 60 paper flashcards to study in preparation for a test on the vocabulary next class. No attempt was made to measure students' flashcard usage outside class. At the beginning of the next class, the paper flashcard post-test was administered. The paper flashcards delayed post-test was given exactly three weeks after the paper flashcards immediate post-test.

3.6. Calculating vocabulary gains

Table 5 shows the mean pre-test (paper and digital tests) scores for each proficiency group.

Table 5. Mean pre-test scores by proficiency level			
	basic	**intermediate**	**advanced**
Mean pre-test scores	1.42	3.72	8.12

The mean pre-test scores (t(212) = 7.71, *p*=.00) were significantly higher for the advanced level group, (*M*=8.12, *SD*=4.78), than for intermediate level group, (*M* = 3.72, *SD* =3.81), who in turn had significantly higher, (t(154)=4.53, *p*=.00), pre-test scores than the basic level students, (*M* =1.42, *SD*=1.69).

Thus, there was more room for improvement for basic students, compared to intermediate and advanced, and for intermediate participants compared to the advanced. Using the raw scores to measure vocabulary gains would lead to inflated gain scores at lower levels of proficiency. To correct for this, relative gain scores (Horst, Cobb & Meara, 1998) were used. This measure considers individual differences in starting positions and is calculated using the following formula:

Post-test score – Pre-test score

Highest possible score (30) – Pre-test score

The maximum possible relative gain score is 1.0. This represents when a participant gets the maximum score (30 for this study) in the post-test, irrespective of the pre-test result. Two relative gain scores were ascertained for each participant, one for the paper flashcards treatment, and the other for digital flashcards.

4. Results

Two, two-way mixed analyses of variance (ANOVA) were conducted to evaluate the effect of English proficiency and Study Mode on vocabulary gains. The dependent variables were Immediate Vocabulary Gain and Delayed Vocabulary Gain both measured from 0.00 to 1.00. The within-subjects factor was Study Mode, with two levels (Quizlet and paper flashcards). The between-subjects factor was English Proficiency with three levels (Basic, Intermediate, and Advanced).

4.1. Within-subjects main effect: study mode

For the sample as a whole, there was a significant main effect of Study Mode on Immediate Relative Vocabulary Gain, *F*(1,136) = 12.87, *p=.00*, η p 2=.09. Digital flashcard Immediate Gains (*M* = .57, *SD* = .23) were significantly higher than for paper flashcards (*M* = .51, *SD* = .28).
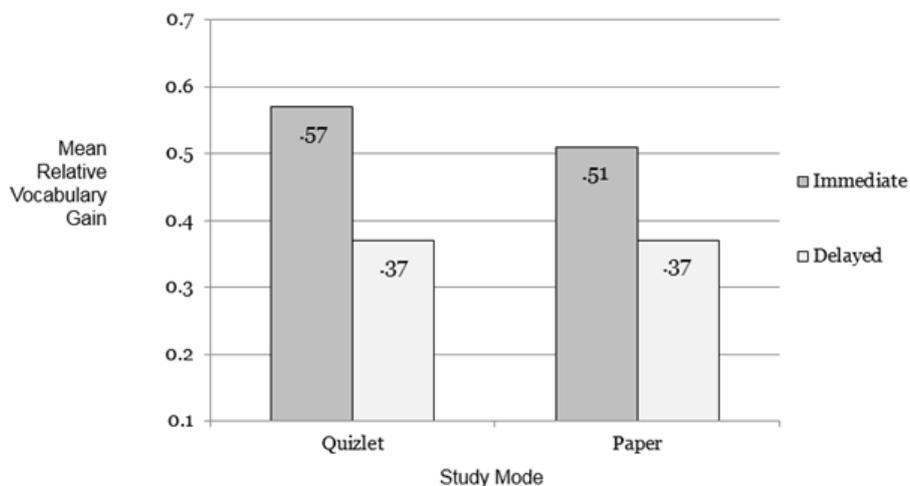

Figure 1. Immediate and Delayed Relative Mean Vocabulary Gains for Quizlet and Paper Flashcards.

However, there was no significant main effect of Study Mode on Delayed Relative Vocabulary Gain. Digital flashcard Delayed Gains (*M* = .37, *SD* = .23) were not

significantly higher than for paper flashcards (*M* = .37, *SD* = .27). This is can be seen in Figure 1.

4.2. Between-subjects main effect: English proficiency

There was a significant main effect of English Proficiency on Immediate Vocabulary Gain, $F(2,136)$ = 26.48, *p*=.00, η p 2=.28. Averaging for Study Mode, the advanced group scored significantly higher immediate gains (*M*= .66, *SD*= .18) than the intermediate group (*M*= .50, *SD*= .22), *t*(105) = 4.28, *p*= .00, who in turn did significantly better than the basic group (*M*=.37, *SD*= .03), *t*(76) = 2.86, *p*=.005.

There was also a significant main effect of English Proficiency on Delayed Vocabulary Gain, $F(2,136)$ = 30.61, *p*=.00, η p 2=.31. Averaging for Study Mode, the advanced group scored significantly higher delayed gains (*M*= .48, *SD*= .19) than the intermediate group (*M*= .36, *SD*= .20), *t*(105) = 3.28, *p*=.00, who in turn did significantly better than the basic group (*M*= .17, *SD*= .13), *t*(76) = 4.53, *p*= .00. These results can be seen in Figure 2.
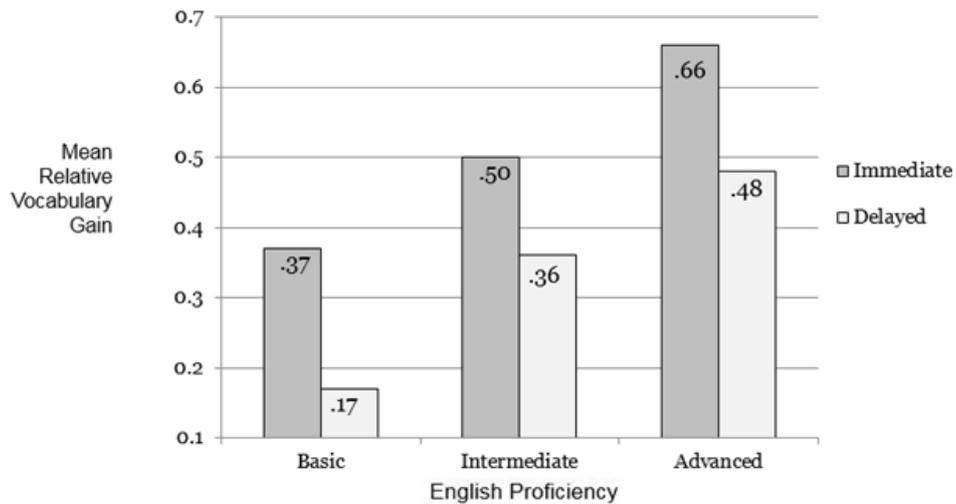

Figure 2. Immediate and Delayed Relative Mean Vocabulary Gains for at different levels of English Proficiency.

4.3. Interaction effect: study mode and English proficiency

There was a significant interaction effect between the Study Mode and English Proficiency both for Immediate Relative Vocabulary Gains, $F(2,136)$ = 4.72, *p*=.01, η p 2 =.065, and for Delayed Gains $F(2,136)$ = 8.42, *p*=<.00, η p 2 =.11. This indicates that the Immediate and Delayed Relative Vocabulary Gain scores differed according to Study Mode and also the level of English Proficiency. In other words, using digital or paper flashcards affected each proficiency level differently. To break this down, multiple comparisons were calculated at each level of English ability for both immediate and delayed gains.

Basic level participants achieved significantly higher Immediate Vocabulary Gains using Quizlet (*M*= .43, *SD*= .21) than when using paper flashcards (*M*= .30, *SD*= .19), *t*(136) = 5.85, *p*= .00. Intermediate level participants also, on average, achieved significantly higher immediate gains using Quizlet (*M*= .55, *SD*= .22) than when using paper flashcards (*M*= .45, *SD*= .27), *t*(136) = 2.9 , *p*=.00. However, the immediate gain results suggest that there was no significant difference between using digital (*M*= .66, *SD*=.21) or paper flashcards (*M*= .67, *SD*= .24), t(136)=-2.45, *p*=.81, for the advanced level group. The relative immediate gains for each study mode and at each proficiency level are shown in Figure 3.

Using paper flashcards, the advanced group had significantly higher immediate vocabulary gains (*M*= .67, *SD*= .24) than the intermediate group (*M*= .45, *SD*= .27), *t*(105)=4.38, *p*=.00, and the intermediate group had significantly higher immediate gains (*M*= .45, *SD*= .27) than the basic group, (*M*= .30, *SD*= .19) *t*(76)=2.72, *p*=.00. Using digital flashcards, the advanced group had significantly

higher immediate vocabulary gains (*M*= .66, *SD*= .21) than the intermediate group (*M*= .55, *SD*= .23) *t*(105)=2.573, *p*=.01. The intermediate group had significantly higher gains (*M*= .55, *SD*= .23) than the basic group, (*M*= .43, *SD*= .21) *t*(76)=2.390, *p*=.02.
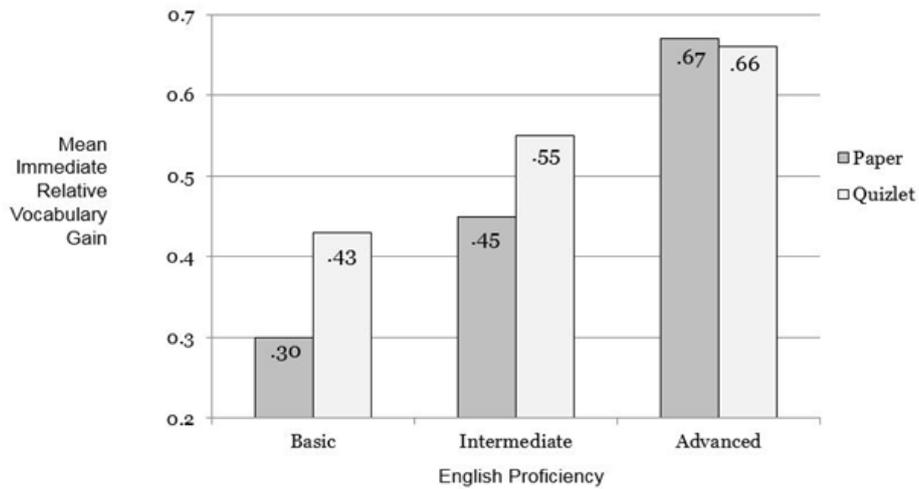

Figure 3. Immediate Relative Vocabulary Gains for Paper and Digital Flashcards for Different Levels of English Proficiency.

The delayed gain scores were significantly higher for basic students for digital flashcards (*M*= .22, *SD*= .16) than for paper flashcards (*M*= .13, *SD*= .13), *t*(31) = 3.69, p =.001. The delayed scores showed no significant difference between using digital or paper flashcards for the intermediate level group. Interestingly, for the advanced group the data showed that participants had significantly higher delayed gain scores with paper flashcards (*M*= .53, *SD*= .23) than with digital flashcards (*M*= .43, *SD*= .21), *t*(60) = 2.97, *p* =.004. The relative delayed gains for each study mode and at each proficiency level are shown in Figure 4.
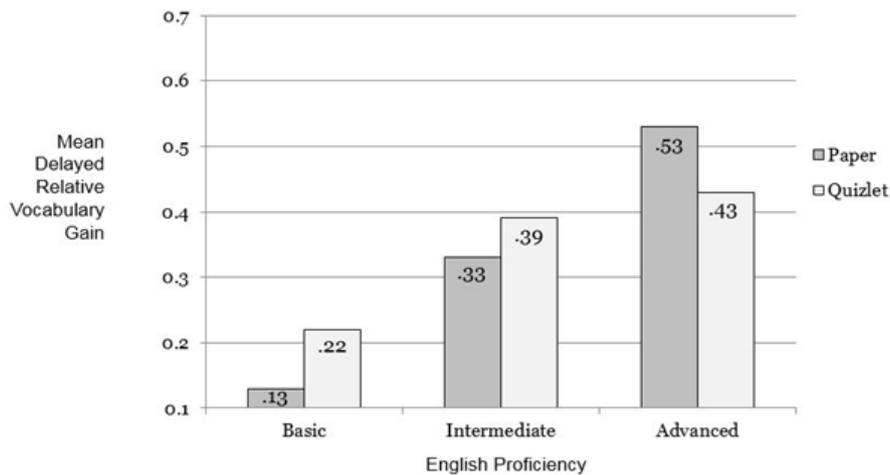

Figure 4. Delayed Relative Vocabulary Gains for Paper and Digital Flashcards for Different Levels of English Proficiency.

Using paper flashcards, the advanced group also had significantly higher delayed vocabulary gains (*M*= .53, *SD*= .23) than the intermediate group (*M*= .33, *SD*= .24) *t*(105)=4.415, *p*=.00, and the intermediate group had significantly higher delayed gains (*M*= .33, *SD*= .24) than the basic group, (*M*= .13, *SD*= .31) *t*(76)=4.254, *p*=.00. Using digital flashcards, the intermediate group had significantly higher delayed gains (*M*= .39, *SD*= .25) than the basic group, (*M*= .22, *SD*= .16) *t*(76)=3.46, *p*=.001. However, there was no significant difference between the advanced group and the intermediate group for delayed gains using digital flashcards.

5. Discussion and conclusions

The results show a significant main effect of proficiency level on both immediate and delayed vocabulary gain scores, indicating that student's proficiency level positively influenced their ability to learn new words, regardless of study mode. The effect of level on relative vocabulary gains is a striking observation. Unlike grammar, it is thought that vocabulary can be learned in any order, irrespective of proficiency level (Lightbrown & Spada, 1999). It would therefore be natural to assume that the words themselves would not account for such a proficiency level effect. A probable explanation is that those participants with better developed aptitudes to learning and higher levels of metacognitive awareness, have become more proficient in English as a result of these qualities. The English proficiency level groupings would therefore also correspond to increasing levels of metacognitive awareness and learner strategy development. Perhaps higher levels of focus, discipline, time management, motivation, confidence, and ability to apply learning strategies, which helped students become more proficient in English, also helped the same students to achieve greater vocabulary gains in this experiment, irrespective of whether they used digital or paper flashcards.

The analysis indicates that digital flashcards were more effective at increasing immediate vocabulary gains than paper flashcards for basic and intermediate-level students. Study mode had no significant effect on immediate vocabulary gains for advanced-level students. This suggests a negative correlation between proficiency, and the comparative superiority of digital flashcards over paper flashcards for L1 – L2 paired associate vocabulary learning. A possible explanation for this result might be that at lower levels the digital flashcards somehow compensated for the lack of metacognitive awareness and learner strategies. Perhaps certain characteristics of the digital application bolstered lower-level participants' lack of such qualities. The variety of activities offered by the Quizlet app, along with the high level of immediate feedback may have helped to boost and sustain the engagement and motivation of lower level students, which paper flashcards could not. The higher levels of control over their study provided by Quizlet due to the delineated nature of the app and access across multiple platforms may also have contributed to maintaining engagement and motivation at lower levels. It seems that by acting as a form of environmental support, the digital flashcards allowed lower level participants to perform to a higher level of proficiency. In contrast, in the experiment, the advanced level students achieved superior results irrespective of the study mode used. It seems that, at least for immediate gains, advanced level students may not require the features of digital flashcards in order to perform well.

The results for delayed vocabulary gains were, however, somewhat different. Again, basic students did significantly better using Quizlet than with the paper flashcards. Unlike the results for immediate gains, there was no significant difference in delayed gains between Quizlet and paper flashcards for the intermediate group. Moreover, the delayed gains for the advanced students were significantly lower using Quizlet than for paper flashcards. The results suggest that the negative correlation between proficiency and the superior effect of Quizlet over paper flashcards is even more pronounced for the delayed gains. In fact, the digital gains were lost significantly more quickly than the paper gains for the advanced level participants.

The higher rate of attrition of the vocabulary gains using Quizlet compared to those of paper flashcards could be attributable to a number of factors. It is possible that the advanced level group did not continue to study using the Quizlet flashcards after the immediate post-tests, while the intermediate and especially the basic group did. Another explanation is that the quality of learning was somehow different between proficiency levels using the Quizlet study cards. In order to discover why advanced gains were more susceptible to attrition, it would be useful to investigate how students used the study modes in the intervening period between the immediate and delayed post-tests. In addition, further research which examines the study behaviour of students outside class and their attitudes towards the two study modes would be helpful to find out how digital flashcards helped lower levels achieve higher learning gains than with paper. Qualitative data collected from paper and digital flashcard users at different

proficiency levels would help to shed light on how attitudes towards the two study modes differs according to level.

The results of this study suggest that digital flashcards help students at lower levels to achieve higher vocabulary gains than when they use paper flashcards. The most advanced group of students in this study however, did equally well with digital and paper flashcards. It seems that the extra functionality provided by the digital platform somehow compensated for lower-level participants' inability to study as effectively as advanced students when using paper flashcards. It seems plausible that lower levels of metacognitive awareness and effective learning strategies associated with lower proficiency level students was cancelled out when using the digital flashcards. This in turn may be due to features of digital flashcards such as greater variety of activities, high level of immediate feedback, increased sense of control and learner autonomy, and the non-linearity of the application. On the basis of these findings, it may be advisable for curriculum designers to consider including digital platforms for L2 vocabulary study for language learners at lower levels of proficiency.

References

Ashcroft, R. J., & Imrie, A. C. (2014). Learning vocabulary with digital flashcards. *JALT2013 Conference Proceedings,* 639-646. Retrieved from http://jalt-publications.org/proceedings/issues/2014-08_2013.1

Baddeley, A. D. (1990). *Human memory: theory and practice.* Hove: Erlbaum.

Cohen, A. D. (1993). *Language learning: insights for learners, teachers, and researchers.* Boston, MA: Heinle & Heinle.

Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly,34*(2), 213. doi: 10.2307/3587951

Cross, D., & James, C. V. (2001). *A practical handbook of language teaching.* London: Longman.

Elgort, I. (2010). Deliberate Learning and Vocabulary Acquisition in a Second Language. *Language Learning,61*(2), 367-413. doi: 10.1111/j.1467-9922.2010.00613.x

Gartner Your Source for Technology Research and Insight. (n.d.). Retrieved March 07, 2017, from http://www.gartner.com/technology/home.jsp.

Hirschel, R., & Fritz, E. (2013). Learning vocabulary: CALL program versus vocabulary notebook. *System,41*(3), 639-653.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond A Clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language,11*, 207-223.

Hughes, A. (2013). *Testing for language teachers.* Cambridge: Cambridge University Press.

Hulstijn, J. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal, and automaticity. In P. J. Robinson (Ed.), *Cognition and second language instruction* (pp. 258-286). Cambridge: Cambridge University Press.

Laufer, B., & Shmueli, K. (1997). Memorizing New Words: Does Teaching Have Anything To Do With It? *RELC Journal,28*(1), 89-108. doi:10.1177/003368829702800106

Lees, D. (2013). A Brief Comparison Of Digital- And Self-Made Word Cards For Vocabulary Learning. *Kwansei Gakuin University Humanities Review,18*, 59-71. Retrieved June 2, 2017, from kwansei.repo.nii.ac.jp.

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.

Nation, I. S., & Webb, S. A. (2011). *Researching and analyzing vocabulary.* Boston, MA: Heinle, Cengage Learning.

Nation, I. (2003). Effective ways of building vocabulary knowledge. *ESL Magazine,* 14-15.

Nation, I. (2005). Language education: Vocabulary. In I. C. Brown (Ed.), *Encyclopaedia of language and linguistics* (2nd ed., Vol. 6, pp. 494-499). Oxford: Elsevier.

Nation, I. (1995). Best practice in vocabulary teaching and learning. *EA Journal,* 7-15. Retrieved March 8, 2017, from http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/1995-Best-practice.pdf.

Nikoopour, J., & Kazemi, A. (2014). Vocabulary Learning through Digitized & Non-digitized Flashcards Delivery. *Procedia - Social and Behavioral Sciences, 98*, 1366-1373.

Puentedura, R. R. (2012, August 23). The SAMR Model: Background and Exemplars. Retrieved March 07, 2017, from http://www.hippasus.com/rrpweblog/archives/2012/08/23/SAMR_BackgroundExemplars.pdf.

Quizlet. (2017). Retrieved March 07, 2017, from https://quizlet.com.

Reinders, H., & White, C. (2011). The theory and practice of technology in materials development and task design. In N. Harwood (Ed.), *English language teaching materials: theory and practice* (pp. 58-80). Cambridge: Cambridge University.

Richards, J. C., & Schmidt, R. W. (2002). *Dictionary of language teaching & applied linguistics.* Harlow: Longman.

Soanes, C. (2010). *The paperback Oxford English dictionary.* Oxford: Oxford Univ. Press.

Suppes, P., & Crothers, E. J. (1967). *Experiments in second-language learning.* New York: Academic Press.

Webb, S. (2007). The Effects of Repetition on Vocabulary Knowledge. *Applied Linguistics,28*(1), 46-65. doi:10.1093/applin/aml048