

# Checking the Effectiveness of Quizlet as a Tool for Vocabulary Learning

## 語彙習得ツールとしてのQuizletの有効性

Blair W.B. Barr, バアル・ブレア

Tamagawa University, Center for English as a Lingua Franca, Japan  
blairbarr@lab.tamagawa.ac.jp

### ABSTRACT

*Inspired by the need to aid low-proficiency English users to increase their knowledge of vocabulary, this preliminary study measures the success of digital-flashcard users by comparing their vocabulary test scores to those of non-users in the same class. Thirty-two, low-level, first year students in Tamagawa University's English as a Lingua Franca program were asked to use Quizlet® to prepare for tests that would recycle fill-in-the-blank contents from the flashcards available on this Web 2.0 application. On average, learners in the class who used gap-fill flashcards with the application scored higher on these tests, and moderately higher on tests with new contents for the same vocabulary. The need for more homogeneous experimentation is acknowledged. In addition, the need for more extrinsic motivation to encourage spaced repetition is considered as means for improving Quizlet-user success on tests of extrapolated knowledge.*

**KEYWORDS:** Quizlet, CALL, Vocabulary, Digital Flashcards

### 1. INTRODUCTION

It is easy to imagine that one of the main focuses of a language program would be grammar, but without words it is impossible to achieve even a basic level of communication. This is not to say that grammatical teaching should cease in the language classroom, but learners I have met often cite a lack of vocabulary and the difficulties encountered studying vocabulary as their main obstacles to learning and using English. One of the aims in my own teaching is to find ways to motivate low-proficiency university students to increase their vocabulary levels.

In this study, learners were encouraged to study with gap-fill flashcards using the digital flashcard application, Quizlet. The aim of the research was to measure the success of users by comparing vocabulary test scores for Quizlet users and non-users of the application. In the process of doing research, Quizlet users were found to consist of two types of learners, kinesthetic and visual, and the following three questions were considered:

1. Are Quizlet users more successful on the vocabulary tests than the non-users?
2. Are kinesthetic users more successful than visual ones?
3. Are Quizlet users more successful than non-users on tests with previously unseen content?

## 2. LITERATURE REVIEW

In my own classes, I make it a practice to have learners use digital flashcards to help them learn and expand their English vocabulary. This is inspired by one consistency in vocabulary learning literature that there is a need for repetition or recycling of language for acquisition to occur (Nation, 2001; Schmitt, 2008). Although the exact number of repetitions is unclear and variable, it is clear that teachers and material writers should develop a longitudinal practice of explicitly repeating the vocabulary learned for the duration of a course (Schmitt, 2008). Neurological studies have shown that time breaks provided by spaced repetition, where language is repeated over increasing time lapses, permits the regeneration of neurochemical substances in our brain that enable long-term recall (Baddeley, 1990).

Many approaches to language learning treat vocabulary learning as incidental (e.g. communicative language teaching and extensive reading). In these cases, words are a part of the communicative or reading process, where students are expected to notice and remember language encountered while focusing on content. This passive approach to building vocabulary, however, has been identified as overly gradual, thus requiring multiple encounters with words before learning can be said to occur (Horst, Cobb, & Maera, 1998; Rott, 1999; Waring & Takaki, 2003). This may be an important part of acquisition, however it is not particularly suitable for the short term objectives of limited-term university-level language training, as such passive noticing and learning would occur too slowly (Schmitt, 2000).

Alternatively, language programs that provide an intentional focus on words may enhance acquisition. Illustrating the advantages of explicit learning, Vidal (2003) found that one month after Spanish learners were exposed to videotaped lectures in English they could recall only words that were central to the theme or garnered explicit focus. In another study, Laufer (2005) found that intentional vocabulary exercises in the classroom not only aided short-term acquisition, but also led to greater retention than incidental learning. Despite such findings, Tang and Nesi (2003) reported that intentional vocabulary training is underused in some classrooms. As a potential solution, Nation (2007) suggests that one quarter of a language program should consist of deliberate vocabulary building, with about a quarter to a third of that time spent using flashcards. In my own classes, I had learners using Quizlet flashcard sets in class with instructions to use them to prepare

for unit vocabulary tests as well.

Regarding the format of flashcards, there are many options available. For one, using L2 definitions or synonyms on the reverse of word cards are options, but these can distract learners with other lexical items that may not be understood well. Another option is to use pictures to stimulate learning with humor and beautiful content, but these may not always be suitable for more abstract target language. Finally, the traditional approach of using L1 translations on the reverse of flashcards has proven to be highly effective for rapid learning (Laufer and Shmueli, 1997). In my own classes, I have set up flashcard sets for each textbook unit. These flashcards stimulate the learners with pictures, but also clarify the meaning of the English words with Japanese translations (Figure 1).

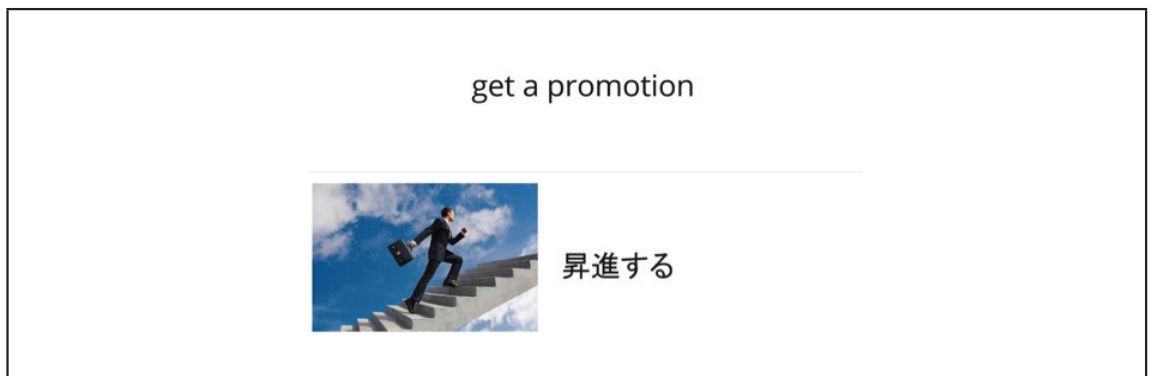


Figure 1. Vocabulary flashcard (screenshot): English words with Japanese translations and pictures.

Repetition of isolated vocabulary alone is certainly helpful for learning spelling and forms (de Groot, 2006), but it is not enough to help learners understand how to use words. In my own classes, I have often found learners had difficulty doing gap activities despite having knowledge of the words through translation. Learners need to be aware of how target words are used in some context (Coxhead, 2008; Horst et al., 1998). Knowing a word through translation is simply not enough. Learners need to know situations where and when words are used. Learners must also know typical collocations, words that typically occur before and after the target words. Learners cannot be aware of such things by studying words in isolation. For this reason, I have provided a second set of word cards that attempts to expose learners to the usage of target words within two or three sentences each. Each flashcard contains one fill-in-the-blank sentence, with the correct form of the target word appearing on the opposite side (Figure 2).

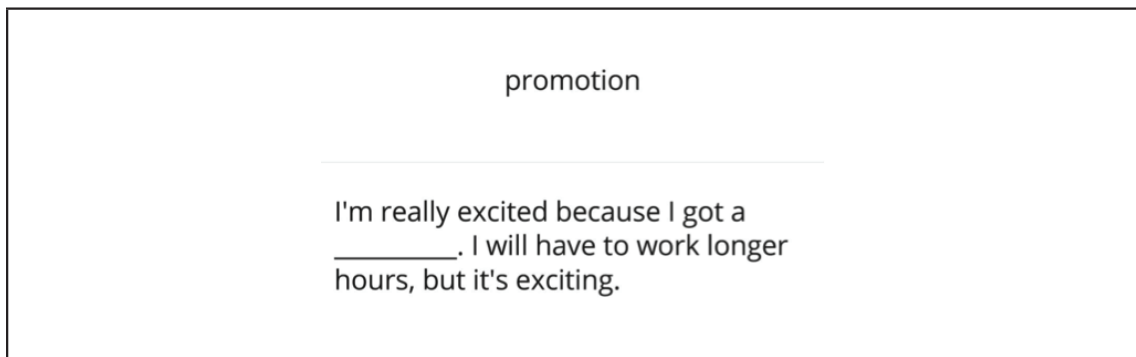


Figure 2. Gap-fill flashcard (screenshot): fill-in-the-blank sentences with target vocabulary.

Quizlet is a digital flashcard application that was used in my classes. Quizlet provides learners with an attractive and easy-to-use interface that can enhance the experience of paper flashcards (Ashcroft & Imrie, 2014). Some of the advantages of Quizlet include the ability to rapidly create flashcards with pictures and audio, the ability to access the application on a computer or smartphone, the ability to automatically rearrange flashcards to avoid serial learning, or memorizing the order, and the ability to interact with the cards using a variety of study and game modes.

This research paper attempted to identify whether or not Quizlet users accessing the gap-fill flashcards were more successful on (1) vocabulary tests recycling content from the cards, and (2) a test with new content using the same vocabulary.

### 3. METHODOLOGY

Thirty-two participants – 15 male and 17 female – in two first-year low-level classes in Tamagawa University's English as a Lingua Franca program were invited to take part in this Quizlet-based study. The participants agreed to permit the monitoring of their Quizlet activities and the anonymous reporting on their vocabulary test results.

For all classes, students were asked to bring a laptop or tablet PC to perform Quizlet activities and take on-line tests, although a few students were permitted to use smartphones. Early in the term, all the students registered identifiable usernames for Quizlet and were invited to join a class page containing private flashcard sets based on the vocabulary from their course book - *Successful Keys to the TOEIC Test: Intro*, with each unit containing 10 new words and 8 phrases based on a related theme. Consequently, as a registered teacher, I could view how and when each user accessed each set of digital flashcards (Figure 3).

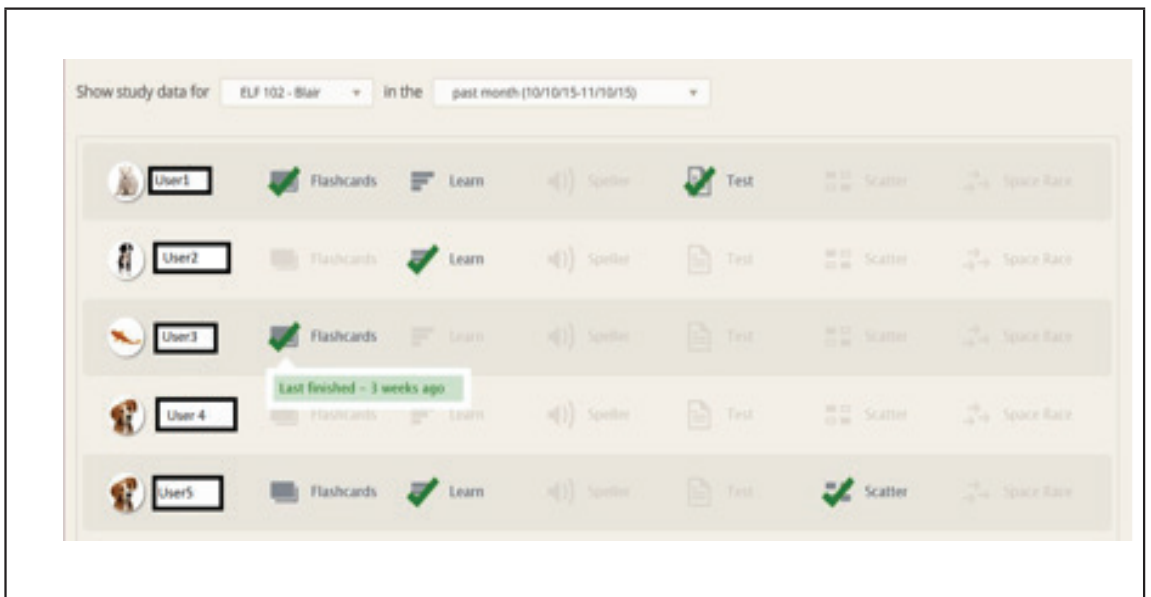


Figure 3. Teacher’s view identifying how and when learners in the course used Quizlet flashcard sets (screenshot – with real usernames obscured).

In class, students were directed to study from two sets of flashcards for each unit. The first set contained pictures and Japanese words on one side and their English equivalents on the reverse (Figure 1). These sets were designed to introduce learners to the vocabulary using their L1 and engaging imagery, while helping learners focus on the base forms of target words. The second set contained 40 to 60 fill-in-the-blank sentences with the correct forms of relevant words completing these sentences on the reverse (Figure 2). The aim of this set was to introduce learners to the words in the context of sentences, with each target word and phrase recurring multiple times in a set. This exposure to multiple usages aims to increase learners’ understanding of how the words are used.

Although all students used these cards at prescribed times in the classroom for about 15 minutes each unit, three groups of users were identified based on their out-of-class usage of the gap-fill flashcard sets to prepare for vocabulary tests. First of all, some participants were classified as *non-users* (e.g., user 4 in Figure 3), who only used Quizlet as directed in class. These participants chose not to prepare for tests using the online sets on their own time. A second group was classified as *visual users* of Quizlet (e.g., user 3), accessing the sets only to view the flashcards or to play the matching game, *Scatter*. For these activities, users are only required to tap or drag the cards on the screen, so the users are simply viewing the language. The final group was classified as *kinesthetic users* (e.g., users 1, 2, & 5), accessing some of the visual learning tools as well as the functions *Learn*, *Test*, *Speller*, and *Space Race* that require participants to physically type correctly-spelled answers. I have classified these users as more

active because they are kinesthetically interacting with the sets through typing while being urged to remember the words, their correct spellings, and their grammatical forms.

For this study, four vocabulary tests were designed for four different textbook units. The test questions were built using a subset of the fill-in-the-blank sentences available in the Quizlet sets. Each test contained 20 gap-fill sentences - 10 multiple-choice and 10 written (Figure 4). The written gap-fill activities were generally more difficult for my past students, so the 10 multiple-choice questions were included for assessment purposes, as they helped increase the overall class average on the gap-fill quiz. Learners were informed to prepare for these scheduled tests. The tests were held within the first ten to fifteen minutes of class, and completed digitally using a *Google Forms* web link, unlocked at the time of the test, and then graded using the *Flubaroo* add-on. The teacher physically and digitally monitored to ensure that participants did not cheat by accessing Quizlet or other applications at the time of the test.

I met my best friend in elementary school, and we have always gotten \_\_\_\_\_ with each other really well.

- problems
- along
- needs
- report

Our manager would like to have a meeting to make a \_\_\_\_\_ about last month's sales.

- deal
- time
- schedule
- report

**Fill in the blank**

I love this class. We have fun, I learn a lot of things, and I get \_\_\_\_\_ with my classmates.

Could I speak to the manager, please? I would like to make a \_\_\_\_\_. My hotel room was so noisy last night. I couldn't sleep.

Figure 4. Screenshot of a vocabulary review test with multiple choice and written gap-fill sentences

The fourth vocabulary test was longer than the first three, as a second part

aimed to test the ability of Quizlet users to extrapolate. Part A was the same as the first three tests and was used for assessing grades. Part B of this test contained 20 additional fill-in-the-blank statements – 10 multiple choice and 10 written – at the end, for a total of 40 questions overall. Part B was designed to expose participants to previously unseen sentences using the same unit vocabulary. This section was a surprise to students, but in order to encourage completion, I did not inform participants until the completion of the test that these additional questions would not be used for assessing their grades due to the surprised nature of the test. This means the participants thought all 40 questions would contribute to their grade, but only the first 20 were part of their evaluation.

In summary, learners' Quizlet activities and the results of the four unit vocabulary tests were used to answer the following questions:

1. Are Quizlet users more successful on the vocabulary tests than non-users?
2. Are kinesthetic users more successful than visual ones?
3. Are Quizlet users more successful than non-users on tests with previously unseen content?

In order to compare success of the two groups in question one, a t-test of independent means was performed to gauge the statistical difference between the means of each group. However, after subdividing the Quizlet users for question two, a one-way ANOVA test was performed to check for statistical relevance between the means of the three groups.

#### 4. RESULTS

At the outset, it is important to note that this was not a controlled experiment. User groups were not predetermined, but rather created by individual study habits. All participants were instructed on how to use Quizlet, and all learners were encouraged to be both visual and kinesthetic users. Any other groups arose from individual preferences while using Quizlet, or poor study habits. It is possible that those participants scored lowly due to other factors. As such, the groups may not be a comparable sample of participants, violating the assumption of homogeneity in the statistical measures.

It is also noteworthy that only some participants were consistently placed in the same user group for each test. The majority of the participants were classified in different groups for different tests. For example, some learners had forgotten about upcoming tests, and failed to prepare with Quizlet on those occasions, becoming non-users. Overall, however, through encouragement and awareness of the benefits of more committed study, the number of kinesthetic users generally increased.

Furthermore, simply because users were classified as non-users of Quizlet, does not mean that they did not study for the vocabulary tests, as some other factors may have helped individuals score well. For one, some students were observed studying the vocabulary directly from their textbooks. Also, a few students used vocabulary notebooks, and they physically wrote out the examples available in Quizlet during the class time for later study. Finally, all students were required to use Quizlet in class, and, as a teacher, I learned to increasingly focus on extrapolating on the difficult fill-in-the-blank sentences with individuals and groups. This last point certainly may have played a role in the increased scores for non-Quizlet users on the last of the four vocabulary tests (Table 1/Figure 5).

Table 1  
*Statistical Distribution of Test Scores for Quizlet Users (Qzlt) and Non-users (NoQ), with T-Test of Independent Means Results*

	Test 1		Test 2		Test 3		Test 4	
	NoQ	Qzlt	NoQ	Qzlt	NoQ	Qzlt	NoQ	Qzlt
Mean	13.57	15.88	9.78	16.74	11.10	17.45	15.38	17.63
Std. Dev.	4.27	3.77	4.38	2.77	3.73	2.58	2.96	2.63
N	14	17	9	23	10	20	13	16
t	$t(29) = 1.60$		$t(30) = 5.40$		$t(28) = 5.47$		$t(27) = 2.16$	
p	$p > 0.05$		$p < 0.01$		$p < 0.01$		$p < 0.05$	

#### 4.1 Quizlet users vs. Non-users

Rather unsurprisingly, the results support the idea that Quizlet users score better on the vocabulary tests with recycled phrases from the cards than those who did not use Quizlet to support their home studies (Table 1). For the first vocabulary test, although the Quizlet users scored higher on average, the difference was not statistically validated. This may partially be explained by the fact that the learners were not entirely certain about their expectations with Quizlet at this stage. However, tests 2, 3 and 4 all resulted in statistically higher scores for the Quizlet users, even when the non-users showed significant improvement on the last of these tests.



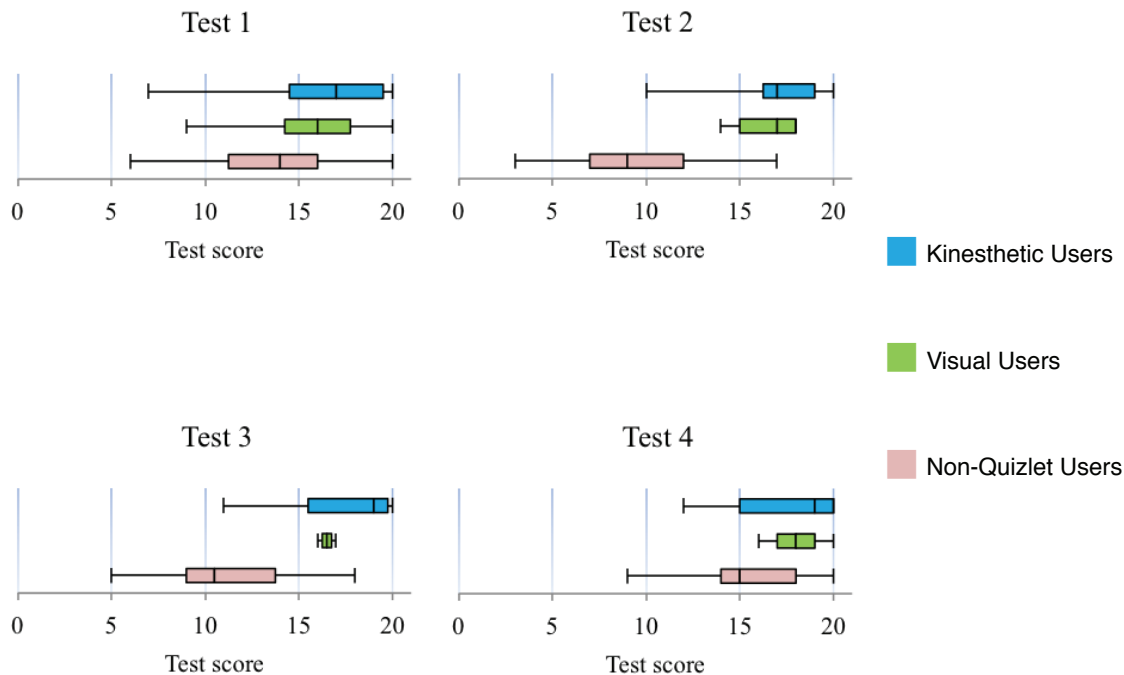


Figure 5. Distribution of test scores for kinesthetic, visual, and non-Quizlet users, with the median represented by the center line, extending outwards to the quartiles and the lowest/highest scores.

#### 4.2 Kinesthetic users vs. Visual learners vs. Non-users

Although the kinesthetic users of Quizlet generally scored slightly higher on average than visual users, the difference between the scores was not supported statistically (Table 2), and this is supported by a box plot of the test score distribution for each group (Figure 5). In fact, the three visual users on test 4 even scored an insignificantly higher average than the multiple kinesthetic users for the same test. Overall, the one-way ANOVA test was most significant for vocabulary tests two and three (Table 2), with most of the difference explained by the lower scoring non-users (Figure 5). Test one may have been lower overall due to confused expectations of Quizlet users, and the higher scores on test 4 may have been influenced by increased teacher support in the classroom. There may be some benefits to using the kinesthetic tools rather than just visual ones in Quizlet, because kinesthetic users usually scored the higher test average. Nonetheless, a larger sample of users may be required to statistically support this assertion. It is also important to remember that the number of visual users was small and decreasing because this group was unintended. The original aim was to encourage all participants to use more Quizlet tools, thus creating more kinesthetic users. However, some users had their individual preferences in practice.

Table 2

Statistical Distribution of Test Scores for Passive (Psv) and Active (Actv) Users, with one-way ANOVA results

	Test 1			Test 2			Test 3			Test 4		
	NoQ	Vis	Kin	NoQ	Vis	Kin	NoQ	Vis	Kin	NoQ	Vis	Kin
Mean	13.57	15.80	16.00	9.78	16.40	16.83	11.10	16.60	17.56	15.38	18.00	17.54
Std. Dev.	4.27	3.3	4.62	4.38	1.82	3.01	3.73	0.71	2.71	2.96	2.00	2.82
N	14	10	7	9	5	18	10	2	18	13	3	13
F	$F(2,28) = 1.24$			$F(2,29) = 14.18$			$F(2,27) = 14.63$			$F(2,26) = 2.28$		
p	$p > 0.05$			$p < 0.01$			$p < 0.01$			$p > 0.05$		

### 4.3 Extrapolation Test

The second part of test number 4 presented the participants with 20 additional gap-fill sentences that were not previously seen in the Quizlet cards. The aim was to determine if Quizlet users were more successful on this test than non-users, thus showing greater capability to extrapolate on the knowledge of the vocabulary they had learned in context. Although, the three groups were identified (Figure 6), I decided to return to the original comparison between Quizlet users (both kinesthetic and visual) and non-users (Table 3) because there were only three visual users for this test.

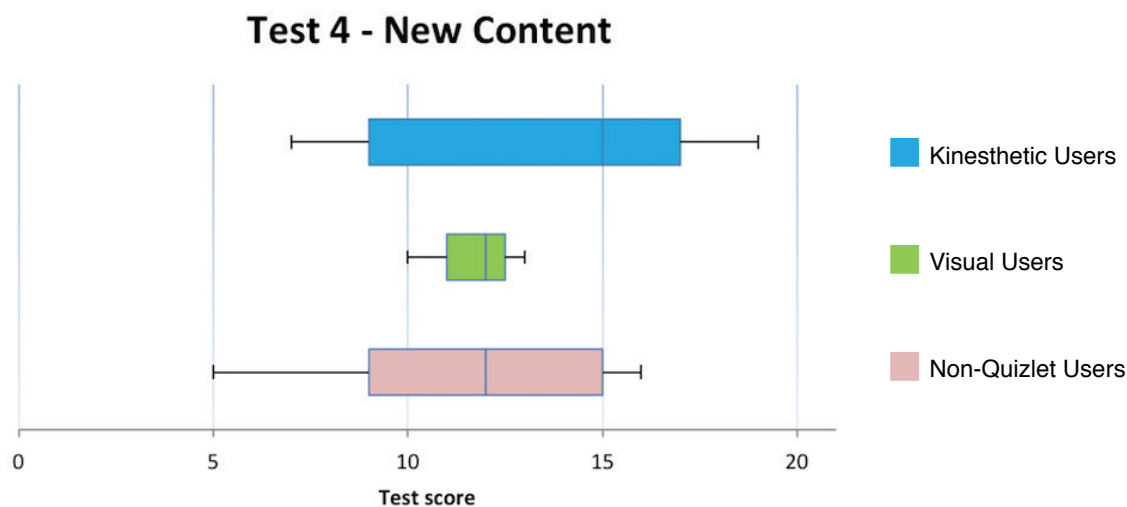


Figure 6. Distribution of active, passive, and non-Quizlet users' test scores on Test 4 with new content, with the median represented by the center line, extending outwards to the quartiles and the lowest/highest scores at the end of the error bars.

Although Quizlet users scored over a point more on average on the vocabulary test with new sentences than non-Quizlet users, the difference between these scores was not supported statistically (Table 3). Again, a larger sample size may verify this difference. Although this one test could not conclusively support the idea that Quizlet users were better prepared to extrapolate on their vocabulary knowledge, the higher average is promising.

Table 3  
*Statistical Distribution of Test Scores for Quizlet Users and Non-users for the Vocabulary Test with New Content, with T-Test of Independent Means*

	No Quizlet	Quizlet
Average	11.77	13.31
Std. Dev.	3.92	3.96
N	13	16
t	$t(27) = 1.05$	
p	$p > 0.05$	

## 5. DISCUSSION

Quizlet users were unsurprisingly successful at recycling their knowledge of the known contents on tests, and their ability to extrapolate better with new content than non-users shows some promising results. However, it is important to remember that all of these results were not comparing two homogeneous groups. In the future, I would try to redesign the experiment to see how the test scores of users of traditional word translation flashcards compare to those using the gap-fill flashcards. Also, I would like to try a controlled experiment with two homogeneous groups, collaborating with a teacher of a group of similar level, but using traditional word cards.

One major problem I found with the approach used in this study is that I failed to motivate continuous vocabulary development. The students generally failed at performing spaced repetition, largely choosing to cram the night before or morning of tests. In addition, students would go on to ignore sets once the unit tests were completed, failing to reinforce the language learned throughout the course.

One option to encourage the increased use of spaced repetition is to have conglomerated sets of flashcards that learners could use to prepare for an end-of-term examination as well. Intuitions about the forms, spelling, and pronunciation of words could improve with the smaller unit sets, and they could prepare themselves

to handle the larger, conglomerated sets (Crothers & Suppes, 1967). Quizlet is well designed for this purpose with a function that permits users to combine sets. In order to provide learners with an incentive to use these sets regularly, a credit system could be devised as a form of extrinsic motivation. This could be in the form of homework completion grades, or, in the case of the ELF program, by providing additional word count credits for the extensive reading requirements in the course. In this way, it is hoped that learners would use some of their time to review vocabulary in context through spaced repetition.

## 6. CONCLUSION

Without a doubt, many of my students were visibly engaged by Quizlet in the classroom. Additionally, it provided learners a way to learn independently on their computers and smartphones. Unsurprisingly, the results of this study demonstrated that learners using Quizlet performed better than non-users on tests recycling contents from gap-fill cards. Although results failed to confirm that kinesthetic users of the application could outperform visual ones, Quizlet users overall could perform a little better on tests with new content. The trends are promising. In the future, I will be implementing more strategies to extrinsically motivate learners to do spaced repetition activities, and looking to compare homogeneous groups. Ultimately, my hope is that these activities will not only play a role in improving learners' test scores, but also expanding their active vocabularies, their confidence, and their enjoyment studying and using the language.

## REFERENCES

- Ashcroft, R. J., & Imrie, A. C. (2014). Learning vocabulary with digital flashcards. In N. Sonda & A. Krause (Eds.), *JALT2013 Conference Proceedings*. Tokyo: JALT.
- Baddeley, A. (1990). *Human Memory*. London: Lawrence Erlbaum Associates.
- Coxhead, A. (2008). Phraseology and English for academic purposes. In F. Meunier & S. Granger (Eds.), *Phraseology in language learning and teaching* (pp. 149-161). Amsterdam: John Benjamins.
- Crothers, E., & Suppes, P. (1967). *Experiments in Second-Language Learning*. New York: Academic Press.
- de Groot, A. M. B. (2006). Effects of stimulus characteristics and background music on foreign language vocabulary learning and forgetting. *Language Learning*, 56(3), 463-506.
- Horst, M., Cobb, T., & Meara, P. (1998). Beyond a Clockwork Orange: acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11(2), 207-223.
- Laufer, B. (2005). Focus on form in second language vocabulary learning.

*EUROSLA Yearbook*, 5, 223-250.

- Laufer, B., & Shmueli, K. (1997). Memorizing new words: Does teaching have anything to do with it? *RELC Journal*, 28(1), 89-108.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. (2007). The four strands. *Innovation in Language Learning and Teaching*, 1(1), 1-12.
- Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition through reading. *Studies in Second Language Acquisition*, 21, 589-619.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329-363.
- Tang, E., & Nesi, H. (2003). Teaching vocabulary in two Chinese classrooms: Schoolchildren's exposure to English words in Hong Kong and Guangzhou. *Language Teaching Research*, 7(1), 65-97.
- Vidal, K. (2003). Academic listening: A source of vocabulary acquisition? *Applied Linguistics*, 24(1), 56-89.
- Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, 15, 1-27.