Using an Online Vocabulary Memorization Tool versus Traditional Vocabulary Exercises[*]

Arif BAKLA[**]

Ahmet ÇEKIÇ[***]

Abstract

This study was conducted to reveal what Memrise, an online vocabulary study tool, can offer to upper-intermediate EFL learners compared to traditional vocabulary exercises in L2 vocabulary learning. Two groups of upper-intermediate learners (N=80) were randomly assigned to the experimental group and the control group and were given the Vocabulary Knowledge Scale, VKS for short, as the pre-test and post-test. The participants in both groups were exposed to the target vocabulary items in the same reading text. While those in the experimental group created list of target vocabulary items collaboratively in Memrise and then studied the sets individually, the learners in the control group did traditional vocabulary exercises. The results of the post-tests indicated that there was a significant difference between the control group and the experimental group in favor of the experimental group. The researchers discuss possible pedagogical implications of this significant finding for EFL vocabulary instruction.

Keywords: L2 vocabulary learning, online vocabulary memorization tool, Memrise, mnemonic devices

Geleneksel Kelime Alıştırmalarına Karşı Çevrim içi Kelime Aracının Kullanımı

Öz

Bu çalışma sınıf içi geleneksel kelime öğretim alıştırmalarının yerine çevrim içi bir kelime çalışma yazılımının kelime öğretimi sürecine dâhil edilmesinin etkilerini belirlemek amacıyla yapılmıştır. Çalışmada, iki grup orta-üst seviyede yabancı dil olarak İngilizce öğrenen Türk öğrenci (n=80) rastgele deney ve kontrol gruplarına atanmıştır. Ön-test ve son-test olarak KBÖ (Kelime Bilgi Ölçeği) verilmiştir. Ön testten sonra, her iki grup hedef kelimeleri içeren yazılı metni okumuştur. Sonrasında, kontrol grubundaki öğrenciler geleneksel sınıf içi kelime alıştırmaları yaparken, deney grubundaki öğrenciler işbirliği içinde çevrim içi kelime öğrenme yazılımı olan Memrise'yi kullanarak hedef kelimelerden liste oluşturmuştur. Son test sonuçları deney grubu lehinde anlamlı fark olduğunu göstermiştir. Bu bulgunun İngilizce kelime öğretimi açısından olası pedagojik sonuçları araştırmacılar tarafından tartışılmıştır.

Anahtar Kelimeler: ikinci dilde kelime öğrenimi, çevrim içi kelime ezberleme aracı, Memrise, hafıza araçları

[*] An earlier version of this paper was delivered as an oral presentation at the 9th International ELT Research Conference (Interdisciplinary Approaches: Beyond the Borders of ELT Methodology) at Çanakkale 18 Mart University, on May 12-14, 2016.
[**] Asst. Prof. Dr., Cumhuriyet University Faculty of Education, Department of Foreign Languages, Sivas. e-posta: arifbakla@gmail.com
[***] Asst. Prof. Dr., Cumhuriyet University Faculty of Education, Department of Foreign Languages, Sivas. e-posta: acekic@cumhuriyet.edu.tr.

Introduction

In line with the commonly cited quote by Wilkins "without grammar very little can be conveyed, without vocabulary nothing can be conveyed" (Wilkins, 1972, pp. 111-112), McLean, Hogg and Rush (2013) emphasize the importance of vocabulary as the foundation of competence in language. Studies into second language (L2) vocabulary indicate that a vocabulary size of 8000 to 9000 word families is required for reading and 6000 to 7000 for listening (Nation, 2006). Although the first 2000 word families can be taught explicitly, Schmitt and Schmitt (2012) state that "there clearly needs to be a focus on vocabulary beyond that covered by the high-frequency, academic and technical categories" (p. 485). As time allocated for learning new words in L2 classes is limited, it is important to use valuable class time as effectively as possible and encourage vocabulary study outside the classroom. There are two research-supported ways learners can learn L2 vocabulary outside the classroom: learning incidentally from either through extensive reading (Saragi, Nation & Meister, 1978; Waring & Takaki, 2003; Pellicer-Sanchez & Schmitt, 2010) or listening to texts (Brown, Waring & Dokaewbua, 2008; Vidal, 2003; 2011) and studying words explicitly by using paired associate learning techniques like wordlists, vocabulary notebooks and flashcards (Nakata, 2008; Dodigovic, 2013).

Learning with flashcards as one of the ways for language-focused learning, which is one of the four strands of vocabulary learning according to Nation (2007), has fallen out of favor in the 1980s with the introduction of more meaning-focused and communicative approaches (Elgort, 2011, p. 367). Criticisms against learning words via flashcards, which is one ways of explicit or deliberate learning of vocabulary, come from two different perspectives, one related with how they are learnt and the other emphasizes ineffectiveness of the results. One of the points of view is that paired associate learning is considered as a relic of behaviorist theory of learning (Hultsjin, 2003 cited in Elgort, 2011) and thus deemed an ineffective way of studying L2 under the reign of more communicative approaches. Second, it is argued that words learnt in this way are not acquired as an active element in the bilingual mental lexicon, thereby making them less automatically retrieved and less likely to be a part of active vocabulary stock of L2 learners.

Although paired associate learning is criticized for being obsolete, Nation and Chung (2009) state that "in terms of efficiency, the most effective deliberate learning of vocabulary involves the use of small word cards with the target word or phrase on one side and the first-language translation on the other" (p. 551). Besides, empirical studies indicate that a large number of words could be learnt through paired associate learning (Imrie, 2014; Hung, 2015). Furthermore, considering the gap between vast amount of vocabulary items to be acquired and time constraints for L2 vocabulary learning in language classes, we cannot solely rely on incidental vocabulary learning through

meaning-focused input or meaning-focused output inside and outside the classroom; there is a need for explicit study of words through language-focused activities to bridge the gap.

As for the second criticism that words learnt in this way do not become a part of active vocabulary, recent studies indicate that deliberately learnt lexical items can also be a part of normal language use (Webb, 2002; Elgort, 2007 cited in Nakata, 2011). In congruency with this claim, the results of a relatively recent study by Elgort (2011) revealed that deliberately learnt vocabulary items can become a part of acquired mental lexicon of L2 learners, which means that explicit and language-focused study of L2 vocabulary is not only necessary but also effective. Accordingly, McLean et al., (2013) argue that both incidental and explicit learning have roles to play in vocabulary learning and one cannot be excluded at the expense of another. Similarly, Lightbown and Spada (1999) argue that unlike grammatical structures, lexical items are especially responsive to deliberate learning as they can be acquired in any order. Therefore, learning L2 vocabulary from flashcards through language-focused study seems to be one of the pedagogical solutions for L2 vocabulary learning. In the view of inadequate body of research into flashcard software, which now offers more innovative learning features like tracking, multimodality and ubiquity, further research is necessary to unearth its possible pedagogical effectiveness over existing traditional vocabulary exercises. To this end, the current study examines the pedagogical value of flashcard software and compares it with traditional in-class vocabulary exercises like gap-fill and sentence writing.

Review of Literature

Since their introduction, digital flashcards have yielded quite a few studies (Nakata, 2008; Dodigovic, 2013; Altiner, 2014; Hirschel & Fritz, 2013). As a relatively recent development, digital flashcard software can be defined as a digital program that enables learners to study L2 lexicon in a paired-associate format (Nakata, 2011); thus, research into its use in EFL classes is still in its infancy. While some studies examined digital flashcards against a set of certain criteria or a model (Nakata, 2011, Ashcroft & Imrie, 2014), other studies examined their effectiveness in L2 vocabulary learning comparing them to paper-based flashcards, word lists or word clouds (Nakata, 2008;Imrie, 2014; Hirschel and Fritz, 2013). Yet another group of studies examined learners' attitudes towards using flashcard websites (Hung, 2015; Chien, 2013; 2015).

Nakata (2008) examined the effectiveness of word lists, cards and computers in L2 vocabulary learning by Japanese learners of EFL. The results revealed no significant difference between all three groups on the posttest, but the computer group performed significantly better on the retention test. Another interesting finding of this study was that time spent studying did not positively correlate with participants' post-test scores for all the three groups.

Ashcroft and Imrie (2014) examined the effectiveness of paper flashcards as opposed to flashcard software (Quizlet) by using the Substitution Augmentation Modification Redefinition (SAMR) model. They found that learners can find definitions of L2 words online and use any one of them, which can save time and effort. This can increase their involvement in vocabulary learning process based on the Involvement Load Hypothesis (Hulstijn & Laufer, 2001). As a result of their systematic comparison of paper versus online flashcard software, Ashcroft and Imrie (2014) concluded that Quizlet does not just substitute paper flashcards but presents a brand new vocabulary learning experience. They further emphasized that a variety of modes offered (flashcards mode, game mode, learn, speller, test, scatter and space race modes) increase multimodality of the materials and learning process, which can augment the chances that we cater for learners' individual differences.

Nakata (2011) examined nine digital flashcard programs across a set of 17 criteria based on research and paired associate learning principle. The researcher pointed out that generally these programs can be used to boost L2 vocabulary learning. The analysis revealed that flashcard creation, multilingual support, adding context, audio and images, variety of activities and active sequencing of learning revealed to be outstanding superiorities of digital flashcards. One interesting finding of the study is the lack of consistent and pedagogically proven guidelines for the development of such programs. Overall, the results of Nakata's analysis point out that flashcard programs stand out as one of the effective and deliberate ways of learning L2 vocabulary.

McLean et al., (2013) studied the effectiveness of using online flashcard site weekly over a period of one year as compared to extensive reading plus deliberate study of words via digital flashcards. The study also included a control group, which was assigned to read texts extensively. The results of their study indicated that studying vocabulary via a flashcards site, which emphasizes paired associate learning and spaced learning principle, can yield significantly higher performances. However, it would be pedagogically more appropriate to assign a mixture of extensive reading and online flashcard study, considering additional benefits of extensive reading. Another study by Altiner (2011) examined Anki as an online flashcard tool to teach 210 words from Coxhead's (2000) academic word list to non-native college students in the US. The results of her study, which used a within-subject design including only 13 students, found that studying words with flashcards improved learners' academic vocabulary knowledge. Yet another researcher, Imrie (2014), compared the effectiveness of using a digital vocabulary-learning tool; namely, Quizlet, as opposed to using paper flashcards and using no such tools in vocabulary learning. The results of the posttest revealed that there were remarkable differences between all the groups, in favor of the Quizlet group, the paper flashcard group and the control group, respectively.

Dodigovic (2013) examined the effectiveness of student-generated and teacher-generated digital flashcards versus paper flashcards and using their favorite learning strategies (e.g. looking up electronic dictionaries etc.) without cards. The results of the study revealed that the learners learnt better when they studied vocabulary using teacher-generated cards or using their favorite strategies. However, there are studies finding that digital flashcards do not lead to better learning performances compared to traditional paper flashcards, word lists etc. A study by Baleghizadeh and Ashoori (2011), which compared the use of word lists and paper flashcards with Iranian elementary EFL learners, found no significant difference between the two groups. In a similar study, Hirschel and Fritz (2013) compared the effects of using vocabulary notebooks versus CALL program with a control group in terms of vocabulary gain and retention. They found no significant difference between the CALL group and the vocabulary notebook group in the posttest. However, the delayed posttest indicated better retention by the CALL group, though it was not statistically significant. Nikoopour and Kazemi (2014) compared the effectiveness of three types of flashcards; namely, paper, online and mobile flashcards. The results of the study indicated that learners in the paper and mobile flashcards group significantly outperformed the online group. The author duly attributed the significant finding to the ubiquity of mobile and paper flashcards rather than modality (digital versus paper). Thus, it is important that learners have consistent access to online or mobile flashcards. In another study, Mansouri (2015) examined the effectiveness of flashcard software versus word clouds, and the results of the posttest indicated that word cloud group outperformed the flashcard group. However, the researcher used PowerPoint to show flashcards, which does not feature the characteristics of digital flashcard programs that allow for collaboration, tracking, spaced learning and systematic scoring and feedback. Therefore, the results of studies by Nikoopour and Kazemi (2014), and Mansouri (2105) should be considered with caution.

Hung (2015) compared vocabulary learning by Taiwanese-L1 students of EFL using digital flashcards under three different task conditions; namely, self-practice, peer-exchange and group-based. The results of the study indicated that the experimental group who worked collaboratively outperformed the two other experimental groups. The researcher also investigated attitudes using Davis' (1989) Technology Acceptance Model to analyze learners' attitudes. The analysis of learner responses indicated that in general all the learners in the study had positive attitudes towards using digital flashcards with learners in the group-based condition having significantly higher scores in all of the three sub-dimensions: usefulness, ease-of-use, and intention-to-use. Chien (2013) also examined the practice and attitudes of L2 learners with regard to making their own flashcards and found that learners tend to choose words to study based on their unfamiliarity with the word rather than by their content or field of study; the results revealed that *Space Race* is perceived as the most useful

activity and that they had difficulty in finding the suitable definition for the word. In another related study, Chien (2015) examined the attitudes of Taiwanese college students towards three popular freely available online vocabulary flashcard websites; namely, Quizlet, Study Stack and Flashcard Exchange based on learner-generated online flashcards, classrooms observations, learning records and interviews. The results revealed that the learners improved their vocabulary knowledge in terms of form and meaning relationship and also had positive attitudes towards learning vocabulary with online flashcards websites and related activities. Lander (2016) examined attitudes of Japanese elementary EFL students with regard to Quizlet upon using it over a period of two 15-week semesters. The learners' responses were examined using a word-mining program and revealed a definite approval of Quizlet.

Current studies compared effectiveness of digital flashcards across different leaner-grouping conditions (Hung, 2015), or as compared with paper flashcards (Imrie, 2014), with extensive reading condition (McLean et al., 2013), with vocabulary notebooks (Hirschel & Fritz, 2013) or with wordlists (Baleghizadeh & Ashoori, 2011) and found controversial results. However, they did not examine their potential to be a more robust alternative to in-class conventional vocabulary exercises like gap-fill and matching. Dodigovic (2013), who cites Godwin-Jones (2010) with regard to the paucity of research evidence about the effectiveness of electronic flashcards, adds that research into student-designed electronic flashcards is even scarcer. To fill in this gap in the literature, this study aims to find out whether collaborative generation of digital flashcard lists by learners themselves can prove more effective and efficient than in-class traditional vocabulary exercises like gap-fill and matching in terms of L2 vocabulary learning. To this end, two groups of upper-intermediate learners were assigned to two different learning conditions and were given the VKS (Vocabulary Knowledge Scale) as the pretest and as the posttest. Through analyzing and interpreting data, this study seeks to answer following research questions:

Q1: Is there a significant difference between the pre-test (VKS) and the post-test (VKS) scores of the both groups?

Q2: Is there a significant difference between the post-test (VKS) scores of the participants in the experimental group and the control group?

Methodology

This quasi-experimental study used a pretest-posttest design and two different learning conditions (Figure 1). Before the treatment, the students were given the VKS and then the students in both groups read the same text in which the target words were presented and answered the same reading comprehension questions. After that, the experimental group used Memrise to study the

target words, while the control group did traditional vocabulary exercises. Finally, the both groups were given the VKS as the posttest.
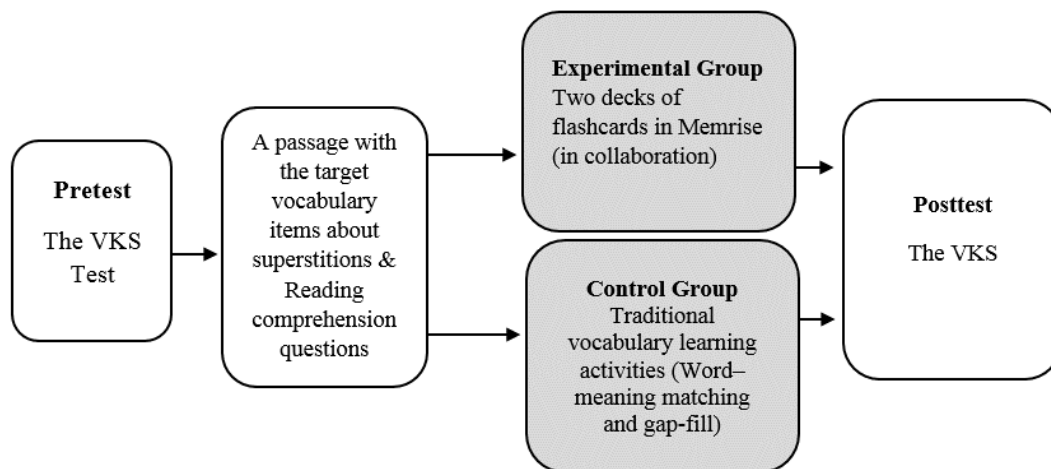

Figure 1. *The overall plan of the study*

Participants

The study group is composed of 80 Turkish-L1 upper-intermediate learners of EFL. Before the study, each student was informed about the study and confidentiality of data. Out of 80 students only 67 were included in the study as 13 students failed to attend either the treatment sessions or the post-test. They had the same level of knowledge with regard to the target words, as revealed by the pretest.

Materials

Flashcard Software (Memrise)

Memrise is an open source (pro version is paid) digital memorization tool which is mostly used for learning vocabulary from a wide range of languages. The software allows its users to create and share their sets of flashcards. Besides, although it is basically designed for individual study, it also enables learners to share their vocabulary sets with the rest of the online community and to customize their own learning.

When evaluated according to 17 criteria (Nakata, 2011), Memrise meets many of the criteria for effective flashcard software. Users are able to add their own mnemonic images audio or text; they are also given a list of mnemonics, which are already uploaded as images. Users can input a variety of types of information into the software, so any type of semantic and grammatical information together with images can be added. It features both presentation and retrieval modes. Once words in the lists are presented with their L2 definitions, retrieval mode starts with receptive recognition test with automatically created multiple-choice items. New words are presented many times based on user's previous performance. In this sense, the software features adaptive

sequencing; in other words, if the word is not known, it is presented again. The software also increases the retrieval effort as it moves on. After receptive recognition, users are asked to write the word itself, which is more difficult as it involves productive recall by giving L2 definition and asking L2 word form. As for the block size, which is defined as the number of cards presented in one learning session, Memrise presents a manageable size block (five words a session). After a session with five words finishes, users receive immediate feedback. Software tracks learners' performance and assists systematic review in line with adaptive sequencing procedure. The users and teacher can track how many words are learnt. In addition, learners can set a goal for a course, which can encourage learner autonomy. As users can share their lists, the software supports collaboration, and teachers can create and enroll learners in their own courses. Thanks to scoring, competition between learners is encouraged the scores in the leader board, which can boost motivation. So, Memrise stands out as an effective way of studying words online either collaboratively or competitively (Figure 2).

Figure 2. *A screenshot of the user interface of Memrise*

The reading text

The students in both groups read an upper-intermediate reading text in a reading based vocabulary textbook titled as the *World of the Words: Vocabulary for College Success* (Richek, 2011). The book is specifically designed to enhance vocabulary learning by upper-intermediate learners and new words are presented in reading texts with pre-reading and post-reading comprehension and vocabulary learning exercises. The reading text we have chosen in this study has 664 running words

with an average readability score of 7.3 (Table 1.), which suggests that the text is suitable for upper-intermediate learners of EFL and conducive to new vocabulary learning.

Table 1. *Readability score produced by various readability tools for the reading passage*

Readability Tool	Score
Flesch-Kincaid Grade Level	4.7
Gunning-Fog Score	6.7
Coleman-Liau Index	12.4
SMOG Index	8
Automated Readability Index	4.7
Average Grade Level	**7.3**

The researchers who have been teaching similar groups of learners for more than 5 years have chosen 18 words from the text based on their corpus analysis of the text's vocabulary profile and frequency of the words in British National Corpus (2007) and Corpus of Contemporary American English (Davies, 2008) by using an online Vocabulary Profiler (https://www.lextutor.ca/vp/)

Table 2. *Frequency bands of the target words according to BNC-COCA-25 (Compleat Lexical Tutor, 2017)*

Word	Frequency Band
Defy	4000 Families
Fidelity	6000 Families
Frantic	5000 Families
Creed	7000 Families
Confidant	5000 Families
Coincide	3000 Families
Crucify	9000 Families
Fiery	6000 Families
Lurk	5000 Families
Coven	14000 Families
Veracity	12000 Families
tongue-in-cheek*	2000 Families
Nonchalant	9000 Families
Delude	6000 Families
Destitute	9000 Families
Discredit	6000 Families
Veritable	9000 Families
Companion	3000 Families

 * (used 342 times as an expression in COCA)

As seen in Table 2, the words are from frequency bands of between 3000-14000 word families. As the participants in this study are upper-intermediate EFL learners, they are considered to have mastered the words from the first 2000 word family. Thus, words from higher frequency bands were chosen as the target words. Only the words in the idiom, "tongue-in-cheek" was from the first 2000 family, but as it is formulaic language rather than taking the constituents frequency into

consideration its frequency in COCA was considered (342 times). So, although the participants know the individual words in it, the researchers expected that they did not know its meaning as an idiom.

Data Collection Tool

To collect data for the study, the VKS (Wesche & Paribakht, 1996) was used. The VKS, which is used as a measure for depth of lexical knowledge, is perhaps one of the most widely used and researched vocabulary measurement tool in vocabulary acquisition research in the literature (Bruton, 2009). As a 5-point self-report scale, the VKS can measure small gains in vocabulary learning and appropriate for incremental nature of vocabulary knowledge development. The developers of the test used test-retest reliability and indicated over .80, which indicates the reliability of the scale. As suggested by its developers (Wesche & Paribakht, 1996), the researchers examined the context the target words were used and set the criteria for correctness at level 4 and 5 was for the target words in this study. The researcher evaluated the VKS papers of the participants separately and the inter-rater reliability was found to be .90. The flowchart below was used to assess the VKS papers.
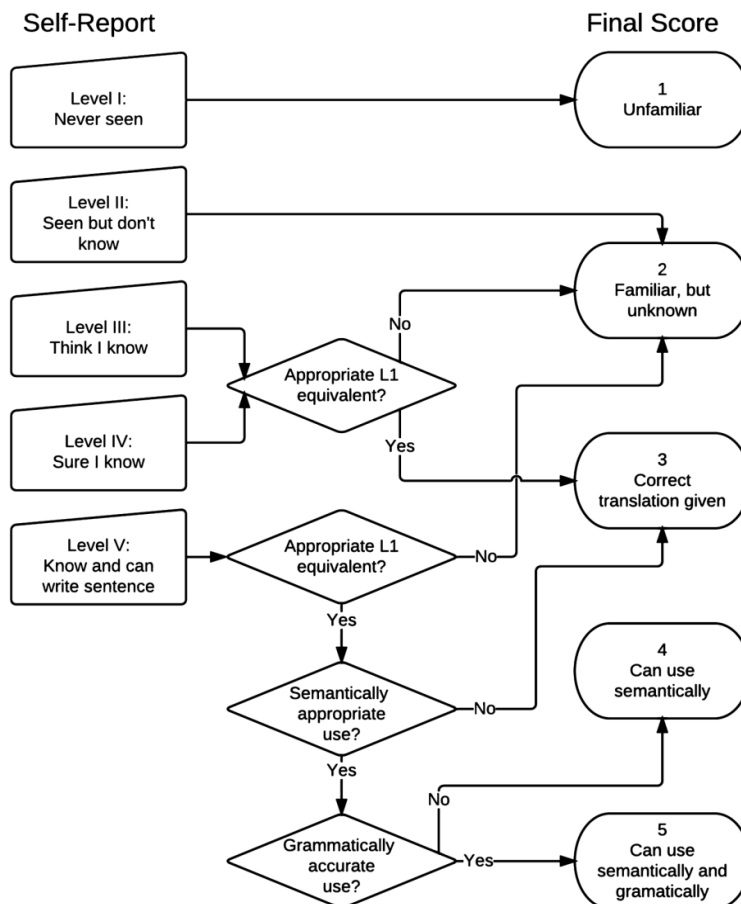

Figure 3. *The flowchart of VKS scoring (Adapted from Wesche & Paribakht, 1996).*

Treatment

Before the treatment process, both groups were given the VKS as the pre-test (the test revealed that there is not a significant difference between the two groups; see results section). The students in both groups did paper-based reading activities. They answered the same reading comprehension questions in class. After the reading comprehension questions, the students in the control group went on to do traditional vocabulary exercises; that is, gap-fill exercises and were asked to write their own sentences using each of the target words. The vocabulary study took 40 minutes. On the other hand, after reading the text and doing the same comprehension exercises, the students in the experimental group did not do traditional vocabulary exercises and studied the words online using *Memrise* digital vocabulary study tool for 40 minutes as well. As the students in the experimental group had already used this digital tool, the researchers thought that it was not necessary to run a training session.

The first researcher enrolled all the students in the experimental group in a *Memrise* class and divided the class into five groups. The groups composed of seven students and each were assigned different tasks to be completed collaboratively. The students referred back to the text to find out the meaning of the bold-faced target words in the context. The first group was responsible for creating a definition match set for target words with short English definitions and added part of speech information. They looked up the words in online version of Cambridge Dictionary (Cambridge Dictionary, 2017). The second group used another online bilingual dictionary (Turkish to English and English to Turkish) (Tureng, 2017) to create a set with L1 equivalents. The third group added synonyms for L2 definition matching and L1 equivalent matching set. They used an online thesaurus (Thesaurus.com, 2017) to complete the task. The fourth group wrote sample sentences for the L1 equivalent set (code mixing in the form of Turkish equivalent of the target word in an English sentence) and for English definition matching set. Meanwhile, the last group added phonetic transcription and audio for the sets. They used text-to-speech software and extracted and copied the link to Memrise (Figure 4). The students worked on all 18 target words, and collaborative tasks took 30 minutes to finish. After that, the students in the experimental group worked individually for 10 minutes and added further relevant mems if they wanted. So, they spent 40 minutes working with the target words.
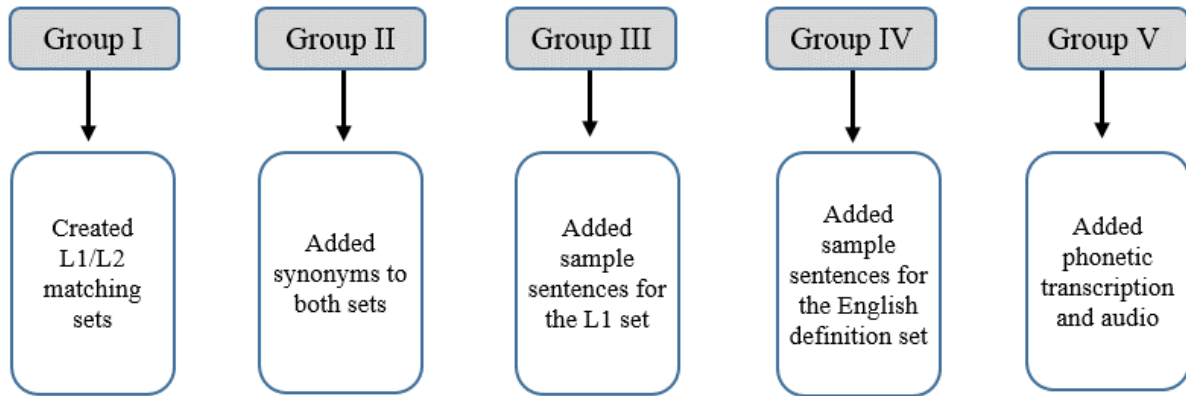
Group I	Group II	Group III	Group IV	Group V
↓	↓	↓	↓	↓
Created L1/L2 matching sets	Added synonyms to both sets	Added sample sentences for the L1 set	Added sample sentences for the English definition set	Added phonetic transcription and audio

Figure 4. *Collaborative work scheme on Memrise*

Results

Data collected from both groups were analyzed for normality of distribution by using Shapiro-Wilk test, which revealed that the dataset was not normally distributed for both groups. Therefore, non-parametric tests were used for the analyses of datasets from the pretest and the posttest as these datasets do not meet the homogeneity of variances assumption.

A Mann-Whitney U test was conducted to determine whether there was a significant difference between the pre-test scores of the experimental (*Memrise*) group and the control group (traditional). The results of the analysis indicated that there was not a significant difference between median scores of the experimental group (MD= 19.50, SIQR= 3.50) and the control group (MD= 21.00, SIQR= 2.25), U= 511,00, p= .529. The experimental group had an average rank of 32.53, while the control group had an average rank of 35.52. This suggests that the groups were approximately equal with respect to depth of knowledge of the target vocabulary items (*semi-interquartile range).

A Wilcoxon sign-ranks tests was carried out to examine statistically significant differences between the scores of the treatment in the pre- and post-VKS test. The post-test scores of the participants in the experimental group (mean rank = 17.00) were significantly higher than their pre-test scores (mean rank= 0.00), (Z= −5.014; p= .001). This finding indicated that the participants in the experimental group boosted their scores in the posttest at significant levels.

To ascertain whether there was a significant difference between the control group's scores on the pretest and posttest, another Wilcoxon sign-ranks tests was carried out. The post-test scores of the participants in the control group (mean rank = 16.39) were significantly higher than their pre-test scores (mean rank= 3.00), (Z= −4.662; p= .001). This finding suggests that the participants in the control group also increased their scores in the posttest at significant levels.

A Mann-Whitney U test was conducted to determine whether there was a significant difference between the post-test scores of the experimental (Memrise) group and the control group

(traditional). The results of the analysis indicated that the experimental group (MD= 32.00, SIQR= 5.75) outperformed the control group (MD= 29.00, SIQR= 4.25), U= 403,00, p= .047. The experimental group had an average rank of 38.65, while the control group had an average rank of 29.21. This suggests that the experimental group outperformed the control group with respect to depth of knowledge of L2 words as measured by the VKS.

Discussion

When the results are examined, it is seen that both groups significantly increased their scores on the VKS posttest compared to their pretest scores, indicating that the flashcard software and traditional vocabulary exercises led to significant learning. This significant result can be interpreted to mean that both treatments are effective in increasing depth of vocabulary knowledge as measured by the VKS. This finding is compatible with the results of previous studies indicating the effectiveness of digital flashcards in L2 vocabulary learning and explicit vocabulary learning through in-class gap-fill exercises (Dodigovic, 2013; Hirschel & Fritz, 2013; Nikoopour & Kazemi, 2014).

The second research question which asked if there was a significant difference between the two treatments in terms of depth of lexical knowledge is also answered positively based on the significant difference in favor of the experimental group on the VKS given as the posttest. The VKS requires learners not only to develop form-meaning associations but also to use words productively in grammatically and semantically appropriate contexts. This significant finding in favor of the deliberate vocabulary learning is congruent with the findings of other studies (Webb, 2002; Elgort, 2007 cited in Nakata 2011; Elgort, 2011), which found that deliberately studied words through paired associate learning can become part of active vocabulary knowledge. This result points to the pedagogical value of digital flashcards for increasing not only receptive lexical knowledge but also for the acquisition of productive knowledge; in other words, grammatical behavior and semantic intricacies of target lexical items can be learnt using flashcard software as it can provide contextualized examples of target vocabulary items (Nagy, Herman, & Anderson, 1985).

The superior performance of the experimental group can be attributed to certain features of Memrise. The participants in the experimental group added ready-made mnemonics stored in the software by previous users or they uploaded their own images. Besides, the learners added audio, phonetic information, example sentences and mnemonics that they find relevant to the target words. As body of research into the use of mnemonics revealed, mnemonic devices can significantly increase vocabulary learning (Atkinson, 1975; Sarıçoban & Başıbek, 2012; Köksal & Çekiç, 2014). Learners' adding their own mnemonics can make learning more customized to individual needs.

Besides enhancing individualization of learning, addition of mnemonic devices could have increased vocabulary gain for two other reasons. First, as demonstrated by previous research (Chun & Plass, 1996; Bisson, van Heuven, Conklin, & Tunney, 2015), multimodality can increase the learning of new words by enhancing the input. Besides, the tasks in the experimental group were in line with need, search and evaluation elements of task induced involvement load (Hulstijn & Laufer, 2001) in that learners need to understand the meaning of target words in text so as to answer comprehension questions, look up (search) the word in a dictionary, find the suitable (evaluate possible meanings) meaning of the word in the context and find sample sentences contextualizing the meaning. Thus, in line with previous research (Tu & Su, 2004; Karalık & Merç, 2016) the significant finding of this study can be considered as a new support for the Task-induced Involvement Load Hypothesis.

Memrise features presentation and retrieval modes. The words are presented in groups of 5 in each cycle. After presentation, in our case, with all the information added (L2 short definition, L1 equivalent, sample sentence, synonyms, phonetic transcription/audio), retrieval practice is achieved through a variety of exercises. Research indicates that retrieval practice in which learners are asked to recall or recognize L2 items and their meanings can enhance vocabulary learning more than mere presentation, because it strengthens retrieval routes to memory (e.g., Barcroft, 2007, McNamara & Healy, 1995 cited in Nakata, 2011).

The significantly higher vocabulary gains in the experimental can also be attributed to a variety of activities offered by Memrise as means for retrieval practice. Nakata (2011) mentions that there are four types of retrieval: receptive recognition, productive recognition, receptive recall and productive recall, all of which should be featured in flashcard software. The software initially requires learners to find the correct L2 definition of a given word among four or more choices, which are the definition of other 4 words in the set (receptive recognition). Later on, it requires learners to choose the target word form matching with the meaning given (productive recognition). Next, the definition of a word is given, and the user is required to write the answer (productive recall). Thanks to a variety of retrieval modes *Memrise* offers, learners' cognitive effort to build memory links is enhanced. The experimental condition involves 3 of 4 retrieval practice excluding receptive recall and the level of difficulty is increased incrementally as the learners move one. Therefore, the significant difference can be attributed to the variety of retrieval exercises. Besides, better performance might be the result of incremental increase of retrieval effort. The retrieval effort hypothesis posits that the more difficult retrieval task gets, the more memory is enhanced (Bjork, 1994; Pyc & Rawson, 2009 cited in Nakata, 2011).

Another superiority of flashcard software condition is that it reassures adaptive sequencing (Nakata, 2011) of lexical item through tracking learners' performance. McLean et al., (2013) state

that adaptive sequencing of flashcards involves more frequent practice of incorrectly recalled words till these words are correctly recalled. Memrise provides immediate feedback and keeps track of learner's performance on a specific item. The learner is presented difficult words more frequently and the words correctly answered are shown to be in the long-term memory. Systematic review of lexical items based on learner's performance in previous sessions can enhance regular review of lexical items, which in turn enhance vocabulary gain (Nakata, 2008; McLean et al., 2013). Therefore, adaptive sequencing and systematic review of lexical items could have led to significantly higher vocabulary gains in the experimental group. Another superiority of the experimental condition, which could have contributed to vocabulary gain, is the immediate and individualized feedback provided for each item. Learners receive individualized feedback in a stress-free environment and can progress at their own speed. All in all, the significantly higher performance of the experimental group could be attributed to the characteristics of flashcard software such as multimodality, presentation and retrieval modes, ready-made mnemonics, adaptive sequencing, immediate and individualized feedback and the nature of the collaborative task, which involved need, search and evaluate elements of the task-induced involvement hypothesis.

Conclusion

The results of the study indicate that flashcard software is more conducive to L2 vocabulary learning compared to in-class traditional gap-fill exercises. As this study is the first to make such a comparison, it bears some alternative and new pedagogical implications for L2 vocabulary instruction and learning. One implication is that instead of doing gap-fill exercises in which some students may choose not to do the exercise or do not pay attention to feedback given, learners can be given chances to create their own flashcards using flashcard software. Besides, flashcard software enables learners to customize learning to their individual needs by ignoring words they know, which is not possible in traditional vocabulary exercises. Furthermore, learners can try out as many times as they can without being anxious about what other students in the classroom think of them. In addition, it can be argued that flashcard software featuring tracking, systematic review and adaptive sequencing can be a sound and research-supported way of learning L2 vocabulary.

Thus, flashcard software (online versions or mobile applications) should be integrated into vocabulary instruction; after the software is introduced in-class, students can be assigned to study words outside classroom. Rather than spending valuable class time for traditional vocabulary exercises in which learners cannot receive individualized feedback and study at their own speed, digital flashcard software can yield better results. Time saved by replacing traditional vocabulary exercises in L2 classes, can be used for meaning-focused input or output activities. In this way, L2 lexical items studied via deliberate paired-associate learning using flashcard software outside the

classroom can be quickly revised in the classroom (via mobile applications if possible) and activated through communicative task-based activities. Besides, the pedagogical criteria set by Nakata (2011) as supported by the significant findings of this study can provide guidelines for teachers to choose and incorporate digital flashcard software in their instruction. In addition, digital flashcard software designers can also benefit from the findings and can follow a more pedagogy-driven and research based instructional design principles. In conclusion, the results of this study indicate that flashcard software chosen carefully in line with criteria set by Nakata (2011), can outperform traditional vocabulary exercises and should replace traditional vocabulary exercises as technology becomes an integrative part of learning languages both inside and outside classrooms.

Limitations

This study has revealed a significant vocabulary gain by the learners who studied target words in collaboration using flashcard software compared to in-class traditional gap-fill and sentence production exercises. However, results are to be taken into consideration by considering the length of study and number of participants. The researchers limited data collection with a set of words that appeared in a reading text as they wanted to control for other variables outside the classroom. Besides controlling for students' possible exposure to target words outside formal instruction period, it was more appropriate to test the words with unannounced test immediately after instructional treatments. Therefore, it was not be possible to keep the instructional period very long as it would make it clear for learners that words they study would be tested. Besides, the number of words to be tested would very high to be covered in one test. However, a more comprehensive test including all dimension of word knowledge beyond the VKS like pronunciation, part of speech and productive spelling tests can be included, and the participants could be given a retention tests to see how the two learning conditions compare when it comes to retention in the long term memory. Prospective studies can investigate acquisition of more words for longer periods of time. Finally, the effects of different recycling patterns on L2 vocabulary gain can be investigated by manipulating the frequency and repetition interval.

References

Altiner, C. (2011). Integrating a computer-based flashcard program into academic vocabulary learning. *Graduate Theses and Dissertations.* 10160. http://lib.dr.iastate.edu/etd/10160

Ashcroft, R. J., & Imrie, A. C. (2014). Learning vocabulary with digital flashcards. In N. Sonda & A. Krause (Eds.), *JALT2013 Conference Proceedings, 639-645.* Tokyo: JALT.

Atkinson, R. C., & Raugh, M.R. (1975). An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. *Journal of Experimental Psychology: Human Learning and Memory, 1*(2), 126-133. https://doi.org/10.1037/0278-7393.1.2.126

Baleghizadeh, S. & Ashoori, A. (2011). The impact of two instructional techniques on EFL learners' vocabulary knowledge: Flash cards versus word lists. *MEXTESOL Journal, 35*(2), 70-82.

Barcroft, J. (2007). Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning, 57*(1), 35–56. doi:10.1111/j.1467-9922.2007.00398.x.

Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2015). The role of verbal and pictorial information in multimodal incidental acquisition of foreign language vocabulary. *The Quarterly Journal of Experimental Psychology, 68*(7), 1306-1326. doi:10.1080/17470218.2014.979211

Bjork, R.A. (1994). Memory and meta-memory considerations in the training of human beings.

In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories.*Reading a Foreign Language, 20*, 136-163.

Bruton, A. (2009). The Vocabulary Knowledge Scale: A Critical Analysis. *Language Assessment Quarterly, 6* (4), 288-297, doi: 10.1080/15434300902801909

Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly, 34* (2), 213–238. doi:10.2307/3587951

Cambridge Dictionary. (2017). *Cambridge Dictionary.* Retrieved from http://dictionary.cambridge.org/

Chien, C. W. (2013). Perception and practice of Taiwanese EFL learners' making vocabulary flashcards on Quizlet. In B. M. Nunes& M. McPherson (Eds.), *Proceedings of the IADIS International Conference e-learning2013*, 459-462. Prague, Czech Republic: IADIS. Retrieved from

http://search.ebscohost.com/login.aspx?direct=true&amp;db=eric&amp;AN=ED562311&amp;lang=tr&amp;site=eds-live&amp;authtype=ip,uid

Chien, C.W. (2015). Analysis the effectiveness of three online vocabulary flashcard websites on L2 learners' level of lexical knowledge. *English Language Teaching, 8*(5), 111-121. https://doi.org/10.5539/elt.v8n5p111

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology, 11*(3), 38-63.

Compleat Lexical Tutor. (2017). Retrieved from http://www.lextutor.ca

Chun, D., & Plass, J. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal*, *80*(2), 183–198.

Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 520 million words, 1990-present*. Available online at http://corpus.byu.edu/coca/.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of Information technology. *MIS Quarterly, 13*(3), 319-340. http://dx.doi.org/10.2307/249008

Dodigovic, M. (2013). Vocabulary Learning: An electronic word card study. *Perspectives (TESOL Arabia), 20*(1), 13.

Elgort, I. (2007). *The role of intentional decontextualised learning in second language vocabulary acquisition: Evidence from primed lexical decision tasks with advanced bilinguals* (Unpublished doctoral dissertation). Victoria University of Wellington, Wellington, New Zealand.

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

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

Godwin-Jones, R., A. (2010). Emerging technologies from memory spaces to spacing algorithms: approaches to second language learning. *Language Learning and Technology*. 15 (2), 4 – 11.

Hirschel, R., & Fritz, E. (2013). Learning vocabulary: CALL program versus vocabulary notebook. *System, 41*, 639-653. https://doi.org/10.1016/j.system.2013.07.016

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349–381). Malden, MA: Blackwell. https://doi.org/10.1002/9780470756492.ch12

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language learning*, *51*(3), 539-558. http://dx.doi.org/10.1111/0023-8333.00164

Hung, H.-T. (2015). Intentional vocabulary learning using digital flashcards. *English Language Teaching, 8*(10), 107-112. https://doi.org/10.5539/elt.v8n10p107

Imrie, A. C. (2014). Using Quizlet flashcards to study vocabulary. In M. McLaren, M. Al-Hamly, C. Coombe, P. Davidson, C. Gunn, & S. Troudi (Eds.), *Proceedings of the 19th TESOL Arabia Conference: From KG to College to Career* (pp. 26-35). Dubai: TESOL Arabia Publications.

Karalık, T.& Merç, A. (2016). The effect of task-induced involvement load on incidental vocabulary acquisition. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 13*(35), 77-92.

Köksal, O., & Çekiç, A. (2014) "The Effects of the Mnemonic Keyword Method on 8th Graders' L2 Vocabulary Learning." *Journal of International Scientific Publications.* (*12) 1313-2571.*

Lander, B. (2016). *Quizlet: What the students think--A qualitative data analysis*. Retrieved http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED572208(=tr&site=eds-live&authtype=ip,uid https://doi.org/10.14705/rpnet.2016.eurocall2016.571

Lightbown, P. M., & Spada N. (1999). *How languages are learned*. Oxford: Oxford University Press.

Mansouri, V. (2015). Vocabulary Instruction: Software flashcards versus word clouds. *Advances in Language and Literary Studies, 6*(1), 41-45.

McLean, S., Hogg, N., & Rush, T. W. (2013). Vocabulary learning through an online computerized flashcard site. *JALT CALL Journal, 9*(1), 79-98.

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly, 20*(2), 233-253. https://doi.org/10.2307/747758

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

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

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

Nation, I. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 59-82. https://doi.org/10.3138/cmlr.63.1.59

Nation, P. (2007). The four strands. *Innovation in language learning and teaching, 1*(1), 2-13. https://doi.org/10.2167/illt039.0

Nation, P., & Chung, T. (2009). Teaching and testing vocabulary. In M. H. Long & C. J. Doughty (Eds.), *The Handbook of Language Teaching* (pp. 543-559). West Sussex, UK: Wiley-Blackwell. https://doi.org/10.1002/9781444315783.ch28

McNamara, D.S., & Healy, A.F. (1995). A generation advantage for multiplication skill training and non-word vocabulary acquisition. In A.F. Healy & J.L.E. Bourne (Eds.),

Learning and memory of knowledge and skills: Durability and specificity (pp. 132–169).Thousand Oaks, CA: Sage.

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

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

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language, 22*(1), 31-55.

Richek, M. A. (2011). *World of words: Vocabulary for college success*, Boston, MA: Wadsworth.

Saragi, T., Nation, P.& Meister, G. (1978). Vocabulary learning and reading. *System, 6*, 70-78.

Sarıçoban, A & Başıbek, N. (2012). Mnemonics technique versus context method in teaching vocabulary at upper-intermediate level. *Education and Science, 37*(164) 251-266.

Schmitt, N., & Schmitt, D. (2012). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching, 47*(4), 484-503. https://doi.org/10.1017/S0261444812000018

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two new versions of the Vocabulary Levels Test. *Language Testing, 18*(1), 55-88. https://doi.org/10.1177/026553220101800103

Spiri, J. (2008). Online study of frequency list vocabulary with the Word Champ website. *Reflections on English Language Teaching, 7*(1), 21-36.

Stewart, J., Batty, A.O.B., & Bovee, N. (2012). Comparing multidimensional and continuum models of vocabulary acquisition: an empirical examination of the vocabulary knowledge scale. *TESOL Quarterly*, *46*(4), 695–721. https://doi.org/10.1002/tesq.35

The British National Corpus, Version 3 (BNC XML Edition). (2007). Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/

Thesaurus.com. (2017). Retrieved from http://www.thesaurus.com/

Tu, H. F., & Su, I. R. (2004). *Effects of task-induced involvement on incidental vocabulary* learning in a second language. (Unpublished master's thesis).National Tsing Hua University, Taiwan.

Tureng. (2017). Retrieved from http://tureng.com/tr/turkce-ingilizce

Vidal, K. (2003). Academic listening: A source of vocabulary acquisition*? Applied Linguistics*, *24*(1), 56–86.

Vidal, K. (2011). A comparison of the effects of reading and listening on Incidental vocabulary acquisition. *Language Learning, 61*(10), 219–258

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.

Webb, S.A. (2002). *Investigating the effects of learning tasks on vocabulary knowledge* (Unpublished doctoral dissertation). Victoria University of Wellington, Wellington, New Zealand.

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: depth versus breadth. *Canadian Modern Language Review, 53,* 1–28.

Wilkins, D. A. (1972). *Linguistics in language teaching*. London: Arnold.