Vocabulary Learning with Electronic Flashcards: Teacher Design vs. Student Design

Marina Dodigovic
Department of English, Culture and Communication
Xi'an-Jiaotong Liverpool University, China
Marina.Dodigovic@xjtlu.edu.cn

Abstract
This article describes a technology-supported deliberate vocabulary learning study, involving students enrolled in various university preparation programs in Qatar. In this study, which uses a within-sample design, the students' vocabulary was pre-tested at the beginning of the academic year. Having identified the most useful vocabulary for this group of learners, both the instructor and the students designed interactive flashcards, using an online tool called WordChamp. The learning outcomes were tested under three conditions: 1) learning vocabulary without cards, 2) learning vocabulary with teacher-designed cards and 3) learning vocabulary with self-designed cards. At the end of the semester, the vocabulary test was administered again, establishing not only a statistically significant improvement in target vocabulary learning, but a significant difference in the effectiveness of the three approaches. While both the contrast procedure and learning with teacher-designed cards were similarly effective, learning with self-designed cards was less effective, at a statistically significant level.

Keywords
Vocabulary, Word cards, Electronic flashcards, Deliberate Vocabulary learning, Explicit vocabulary knowledge, Vocabulary test, WordChamp, Flashcard design

Introduction

Research suggests that, in addition to incidental vocabulary learning, which occurs in the context of communication and extensive reading, deliberate vocabulary learning strategies facilitate vocabulary growth (Oxford & Crookall, 1990; Nation 1990; 2006; Schmitt, 2008). Such learning is expected to lead to explicit word knowledge (e.g. R. Ellis, 1997; N. Ellis, 2001), and could be thus said to promote language awareness (Svalberg, 2007). Flashcards are an important tool in deliberate vocabulary learning (Nation, 2006; Oxford & Crookall, 1990), traditionally involving paper or cardboard cards with text, image or numbers on one or both sides (Folse, 2006; Schmitt, 2008). Such cards are usually held up for a brief period of time to help the learner memorize a word or concept. The most common formats in second language vocabulary learning include L2-L1 cards, or cards with the target language word (L2) and their translation (L1), or word-definition cards, with or without examples of use (Oxford & Crookall, 1990).

In the literature (Oxford & Crookall, 1990; Nation 1990; 2006), it is often assumed that word cards would be created by students themselves. In order to accomplish this, the students would first have to look up target words, a procedure conducive to noticing, which is deemed to be a prerequisite to learning (Nation, 2001). Moreover,

the students would have to write down the words and organize in meaningful categories information about word meanings, connotations, pronunciation or context; such an effort, it has been thought, would encourage students to deepen and solidify their knowledge of individual words (Nation, 2001; Folse, 2004; Canado, 2010). According to Nation (2006), writing down information about a word helps one remember it.

However, the depth of processing needed to create one's own cards may not be as conducive to retention as is the "number of word retrievals required" (Folse, 2006, p. 273). This begs the question of whether more effective learning could be achieved by repeatedly using word cards that are not learner-designed.

Unlike traditional cards, electronic word cards require the learner to type in the word during retrieval, thus combining the advantage of writing with that of ready-made cards. They can also manage the learning process, utilizing what is known about the transition from short-term into long-term memory (Nakata, 2008). However, there is little research evidence regarding the effectiveness of such cards (Godwin-Jones, 2010), let alone the effectiveness of student-designed electronic word cards. The classroom experience of the present author with electronic word cards points to a student preference for teacher-designed cards over those they developed themselves. This has led to the research question: whether teacher designed electronic word cards are more conducive to vocabulary learning than student-designed cards. The study at hand addresses the question in a quantitative research context.

Literature review

According to Alderson (2005, p. 88), "language ability is to quite a large extent a function of vocabulary size". While each language skill or the context in which the language is used may require a different size and strength of vocabulary, one constant remains, i.e. word frequency, within the target context, as a measure of word usefulness (Schmitt, 2010; Nation, 2001). It has been suggested that keeping abreast of tasks at an English medium university, in addition to the 2000 most common words, requires the knowledge of 570 otherwise less frequent academic words (Academic Word List or AWL; Coxhead, 2000). Mastering these words in meaning, form, context and function (Nation, 2001; Schmitt, 2010) in pre-university English language classes would empower second language students to succeed in university courses (Coxhead, 2000). Hence the effectiveness of vocabulary learning strategies emerges as an important issue (Nation, 2001; 2008).

Incidental vocabulary learning, supported by extensive reading (e.g. Nassaji, 2003; Schmitt, 2008), requires six to ten encounters with a word, which, with fiction as the basis, is afforded for the 570 academic words (Cobb, 2007). While such encounters can be facilitated by creating additional reading packs (Horst, Cobb, & Nicolae, 2005; Cobb, 2007), there might be cultural constraints regarding the amount of reading expected of students. There are also concerns regarding the effectiveness of incidental vocabulary learning (Won, 2008; Schmitt, 2010). Consequently, incidental learning in some contexts needs to be supplemented by intentional learning (Nation,

2006; Schmitt, 2008). Flashcards are a tool often used for such learning (e.g. Nation, 2006; Oxford & Crookall, 1990). It is claimed that flashcard facilitate recall, i.e. more than simple recognition (Laufer et al., 2004). L2-L1 cards utilize the existing L1 (first language) meaning, already present in the learner's memory, and therefore support the creation of the initial form-meaning link (Schmitt, 2008; Nation, 2001).

On the other hand, cards with examples of use (Oxford & Crookall, 1990) might facilitate the acquisition of a deeper knowledge of a word (Schmitt et al., 2011). Oxford and Crookall (1990) differentiate between decontextualizing, semi-contextualizing and fully contextualizing vocabulary learning strategies. Word cards, along with word lists and dictionaries are deemed to be decontextualizing techniques, so long as L2-L1 word pairs or definitions are used. However, adding examples of use as well as visual or aural imagery would, according to Oxford and Crookall (1990), make the technique semi-contextualizing. This differs from Nation (2001), according to whom even sentence examples are decontextualized, so long as they are removed from the context in which the original message occurred. For the purposes of this study, Oxford and Crookall's (1990) terminology is more useful, allowing for more precision. Consequently, cards based on L2-L1 word pairs, synonyms and definitions are called *decontextualizing* in this text, whereas those with examples of use or other information about the word are called *semi-contextualizing*.

Studies on electronic word cards have been conducted under decontextualizing conditions. Apart from introducing writing into the recall activity, an advantage of such cards over conventional flashcards is their management of the learning process. Thus, a study by Christensen et al. (2007) demonstrates the relative effectiveness of electronic flashcards, using words and definitions only. The program in which they are embedded focuses on the items found difficult by the user in such a way as to take advantage of the spaced learning phenomenon. In spaced learning, rehearsal opportunities are spread over a longer period of time, in gradual increments (Nakata, 2008; Nation, 2006). This has been found to be more effective than massed learning, where rehearsal is restricted to a smaller number of sessions within a shorter period of time (Nation, 2001; Nakata, 2008). Moreover, online word card drills mark the words not successfully recalled, and present them to the learner in gradual time increments (Christensen et al., 2007; Nakata, 2008). In addition, the cards are shuffled to avoid the serial learning effect. Nakata (2008) finds L2-L1 electronic word cards, based on the above spaced learning principles, as effective as physical L2-L1 word cards, while Chen, Hsie and Kinshuk (2008) demonstrate a facilitating effect of mobile phone-supported flashcards on short-term memory retention of vocabulary.

While the latter study includes pictorial annotation, which could be thought of as a semi-contextualizing feature, there are other ways of achieving semi-contextualization. Including examples of use, or other contextual and functional information, such as collocations, in deliberate online vocabulary learning could have a semi-contextualizing effect (Oxford & Crookall, 1990). Such additional information could be based on what is assumed about the vocabulary learning process. Schmitt (2010) believes that learners first remember some elements of form (e.g. the number

of syllables, the initial letters, aspects of pronunciation, such as what other word it rhymes with) and meaning. At a later stage, these may be followed by observations about collocations, register and constraints (Schmitt, 2010). Initial letters of words are already frequently utilized to trigger memory in productive vocabulary testing (Laufer et al., 2004), while Nation (2001; 2008) suggests integrating synonyms, antonyms, associations, collocations and constraints into vocabulary learning. Although care must be taken not to overwhelm the learner with information (Nation, 2001; 2008), each learner could choose from the above options to create meaningful, individualized word cards. In an online learning environment, such a selection could be facilitated by sound-enhanced dictionaries and concordances, features that might add value to noticing as the key to vocabulary learning (Nation, 2001).

Research indicates that active engagement with the material, required in the production of such semi-contextualizing cards, might have added benefits (e.g. Horst, Cobb, & Nicolae, 2005; Nation, 2006). It would require looking up and mentally processing semantic, contextual, functional and grammatical information about a word, i.e. those components that are needed for full mastery of a lexical item (Folse, 2004; Schmitt, 2010). In addition, Nation (2006) believes that writing words down is key to memorization, while Oxford and Crookall (1990) claim that "the copying component…might provide a small kinesthetic benefit for some learners" (p. 12).

As Godwin-Jones (2010) points out, more research into the long-term recall effects of electronic word cards is needed. While flashcards seem to have been a common feature of CALL (Computer Assisted Language Learning) programs, except in the above-mentioned studies, not much research using these has been reported. On the surface, it may seem like much has been written about electronic word cards, as articles containing this keyword started appearing in the CALICO Journal in the early eighties (e.g. Baker, 1984). However, these articles, even when taking a critical stance (e.g. Baker, 1984; Bush & Browne, 2004), do not tap into the theoretical underpinnings for or against the use of word cards, nor do they include studies based primarily on such cards. Thus, the subject of electronic flashcards presents a real niche to be filled with research evidence, the more so as the now ubiquitous mobile technologies increasingly support flashcards applications, such as Anki or Quizlet (Godwin Jones, 2011), to name just a few. The present study seeks to establish whether there is a difference between the effectiveness of teacher-designed and student-designed electronic word cards. It does so using cards in which sentence examples as well as other formal, semantic, functional and contextual clues assume a semi-contextualizing function.

Rationale

The idea for this study emerged from the classroom while activities designed with the online vocabulary teaching program called WordChamp (www.WordChamp.com) were being conducted. One of the many features of this program is a word-card builder, which can be used by students and teachers alike. Its drill program, based on the principles of spaced learning, can be used for vocabulary recall, which requires typing in a word, either with full or approximate precision. A number of words stored

in the system's word bank come with a sound recording, which is played after successful or unsuccessful word retrieval. The card builder also allows the inclusion of other multimedia types. In addition, WordChamp is a Course Management System (CMS), i.e. an online tool enabling instructors to post material and monitor the progress of students in a safe environment.

Working with WordChamp cards, the present author encountered a spontaneous expression of preference, from an entire class, for teacher-designed cards over those of their own design. Since affective factors are known to have an impact on learning (e.g. Sawyer & Ranta, 2001; Ortega, 2010; Krashen, 2004), it seemed of interest to compare the learning outcomes achieved through the use of the two types of cards. Hence the research question: Do students benefit more from teacher-designed semi-contextualized text-based electronic flashcard activities than they do from those they designed themselves?

Methodology

Participants and study design

In this study, the number of participants and the study design were based on the required sample size, using tables (Cohen, 1988). To minimize error and assure reliability at the level of 0.900-0.95 for the desired significance level of 0.5, it seemed that 85-105 participants per group were required. The sample was a convenience sample, as scheduling restrictions on both participants and researchers were considerable. One hundred and two 17-19-year-old students in pre-university foundation programs in Qatar whose schedules allowed for this chose to participate as respondents. All were native speakers of Gulf Arabic, with PBT (Paper Based TOEFL) scores ranging roughly from 350 to 450. Although not randomized, this sample appears to be consistent with the local population of university applicants in Qatar.

Since this number was not large enough to allow for the inclusion of a treatment and a contrast group, without risking a decrease in the statistical power of the tests to be subsequently performed with the data, a within-sample or repeated measures design was chosen, in which the same participants, rather than two different groups, were exposed to different treatments (Al-Seghayer, 2001; Felix, 2008). An advantage of this design is deemed to be increased reliability, since each participant acts as his or her own control "by taking part in each of the conditions" (Felix, 2008, p. 149). This leads to the exclusion of some of the confounding variables which often render between-subjects designs with small groups unreliable.

Participants were exposed to two different types of word card-based treatment (teacher-designed and self-designed semi-contextualizing electronic word cards). In addition, a third activity characterized by the absence of word card-based treatment was introduced to serve as a contrast procedure, in order to rule out possible confounding variables (McDonough & McDonough, 1997). The word card activities in the two card-based procedures were practiced at intervals recommended by Nation (2006), i.e. taking longer breaks between the repetition of the activity (5 minutes,

then half an hour, then three hours etc.). This is based on the phenomenon of spaced learning, which requires a gradual increase in the interval between recalls and is deemed to be more effective than massed learning, i.e. frequent repetition within a shorter time interval (Nakata, 2008, Nation, 2006). In this study, the participants were advised to repeat the activity on a daily basis, using the following increments: one hour, five hours, ten hours.

Within-subjects or repeated measure design has, however, ramifications of its own, consisting largely of the confounding impact of treatment order, be it as carryover effect, i.e. continued effect of the previous treatment, or fatigue, i.e. saturation with a similar task (Algina et al., 2005). In this study, the difficulty of the word group to be studied was another potentially confounding variable. To control for the effect of any interaction between the word groups and the type of condition, i.e. to avoid the coupling of the same condition with the same word sub-list word for all participants, word groups were varied across the three conditions, the order of which was in turn also varied. Thus, there were six possible permutations of condition (1, 2, 3; 1, 3, 2; 2, 1, 3; 2, 3, 1; 3, 1, 2; 3, 2, 1) and word group (e.g. X, Y, Z; Y, Z, X etc.), leading to 36 (6 x 6) combinations (e.g., 1X, 2Y, 3Z; 3X, 2Z, 1Y etc.), which were randomly assigned to participants, as modeled by Folse (2006). This was done to ensure that the results were not affected either by the order of the conditions or by the relative difficulty of word sub-lists.

The independent variable in this study was the condition, while the dependent variable was the learning outcome, as measured by vocabulary tests. Consequently, the activities, focusing on a subsection of the AWL or Academic Word List (Coxhead, 2000), were preceded by a pre-test shortly before the start of the condition sequence and followed by an achievement test a week after each activity and then a final vocabulary test two months after the completion of all of the learning and achievement testing activities, to measure the overall gain.

Instruments

A combination of measurement instruments was used, including the Vocabulary Levels Test (Schmitt, Schmitt, & Clapham, 2001) at the beginning and at the end of the semester, the Versant Test of English (an automated spoken English proficiency test with a significant vocabulary component) at the beginning of the semester and a battery of compatible vocabulary post-tests, used after each condition. The purpose of the initial VLT was to identify the area of vocabulary weakness, i.e. to ascertain that participants had learning needs for academic vocabulary. According to Nation (2006), a result below 80% on any subsection of the VLT indicates a deficiency at that vocabulary level. Accordingly, participants scoring an average below 80% on the AWL section of the VLT, confirming that they have a learning need in respect of academic vocabulary, would qualify as suitable for the sample for this study. The final VLT score was obtained to measure the overall academic vocabulary gain at the completion of the study.

There were three purposes in using the Versant. Firstly, it was used to measure the phonetic coding ability of the participants, and secondly, to get a depth-of-vocabulary measure through extended responses (Nation, 2006). Thirdly, it provided a baseline English proficiency measure, which was subsequently used to explain any possible differences in vocabulary learning outcomes. The Versant was not re-administered at the completion of the study, since the participants were enrolled in intensive English programs, which would have affected the scores as a confounding variable. These programs, however, did not focus on AWL, crucial to this study.

To account for any increment in academic vocabulary gain, similar to Folse's (2006) study, and taking on board some of Schmitt's (2010) concerns, a modified Vocabulary Knowledge Scale (VKS; Paribakht & Wesche, 1993) was used for the post-test. It identified one of the following: I. recognition (0 or 1) or II. recall (0 or 1 or 2), with the highest score being 3 for each word (see Appendix). While the recognition level required participants to supply an English synonym or translation, recall was demonstrated by use of the word in a sentence. The score for recall included a point for the correct lexical context and a point for the correct form (see Paribakht and Wesche, 1993).

This VKS-based instrument seemed more useful than the Word Associates Format (WAF), a test relying on word association, which does not lend itself to reliable interpretation (Schmitt et al., 2011). Similarly, the Schmitt and Zimmermann Scale (Schmitt, 2010), a form of self-assessment, was ruled out because of the limited reliability of self-assessment (Lew et al., 2010).

Learning materials

Both teacher and student-designed cards were to include, and all teacher-designed cards actually did include, semantic, phonetic, contextual, morphological or syntactic clues about each word. An example is shown in Figure 1. In order for students to be able to notice and extract this information, they were trained to use tools such as online dictionaries and concordances. While AWL was identified as the target vocabulary, it was important that the target vocabulary present the same learning burden, i.e. the relative difficulty of word form and function (Nation, 2006; Folse, 2006) to all study participants. One way of controlling this variable across the three conditions was to limit the range of grammatical categories (Folse, 2006). Verbs and adjectives were thus selected as the only grammatical categories for the target vocabulary across the three conditions, because of their medium level of difficulty (Laufer et al., 2004).

Figure 1: Drill with a teacher-designed card

The second criterion of difficulty was that each target word had to be unknown to all study participants. For this reason, common lexical gaps (Kaur & Hegelheimer, 2005) were identified, mirroring a procedure used by Horst, Cobb and Nicolae (2005), where the participants were asked to identify the unfamiliar words in several academic texts. Thirty verbs and adjectives in the AWL category that were not known to any participants were then selected as target vocabulary for the study. They were subsequently divided into three lists of ten words each. Each participant studied each of the three lists, one for each of the conditions. However, there was no fixed association between a word list and a condition.

Procedure

At the beginning of the semester, the Vocabulary Levels Test and the Versant were administered during the first week of classes, followed by the identification of the target vocabulary consisting of thirty academic words, all unfamiliar to each of the participants. Students were then introduced to the WordChamp cards. Subsequently, the three conditions were introduced, as described in the design section, i.e. by varying the order of both word groups and conditions, leading to 36 different combinations, randomly assigned to students, with approximately three students following the same path at any one time. Each student received three emails, one for each condition, explaining the exact nature of the task at hand. When designing their own cards, the students had no access to teacher-designed cards and vice versa.

While using teacher-designed cards involves logging into a specific online class on WordChamp and simply following the homework link to activate the card drill exercise, designing cards from scratch is more complex. Firstly, the user has a choice of several card types, such as definition or question-answer card. Whereas the

selected template can then be easily filled with content, the choice of information to include is up to the user. Hence the quality of a card largely depends on what the user has decided to associate with a word. If the learner has, for example, included a definition which is too vague or too broad and does not include any mnemonics, that learner may not benefit from using such a card, once it has been created. The participants were encouraged to use online monolingual learner dictionaries and concordances to locate and include useful information about a word, including its meaning, form and possible contexts for use.

The contrast procedure was similar to the two treatment procedures in that ten new words were first introduced within the context of a reading passage, in which the most of the rest of the vocabulary was familiar to the students. Unlike in the card-based treatment procedures, in the contrast procedure the students were asked to look up the new vocabulary and learn it. While they were already acquainted with electronic dictionaries and concordances, and it was recommended that they use these for learning purposes, it was left up to them to use the method of studying that worked best for each of them. Only the participants meant to be working with electronic cards were allowed to log in to the WordChamp resources created for them for the duration of the task. Thus, students engaging in the contrast procedure were unable to use teacher-designed cards or create their own.

For the purpose of the second condition, the students were asked to make spaced use for a week of the WordChamp cards produced by the teacher. WordChamp logs (see Figure 2) were used to track progress and make sure that all of the students completed the exercise. The third condition required the students to design their own flashcards. Earlier on, they had a training session in which they were shown how to create cards in WordChamp, followed by practice under the supervision of an instructor. For the purposes of the third condition, card design was assigned for homework, after which the students again had a week to engage in spaced learning using these cards. Each of the conditions was followed by a post-test for the respective word group, within a week. Finally, another version of the VLT was administered again, with the focus on the AWL section.

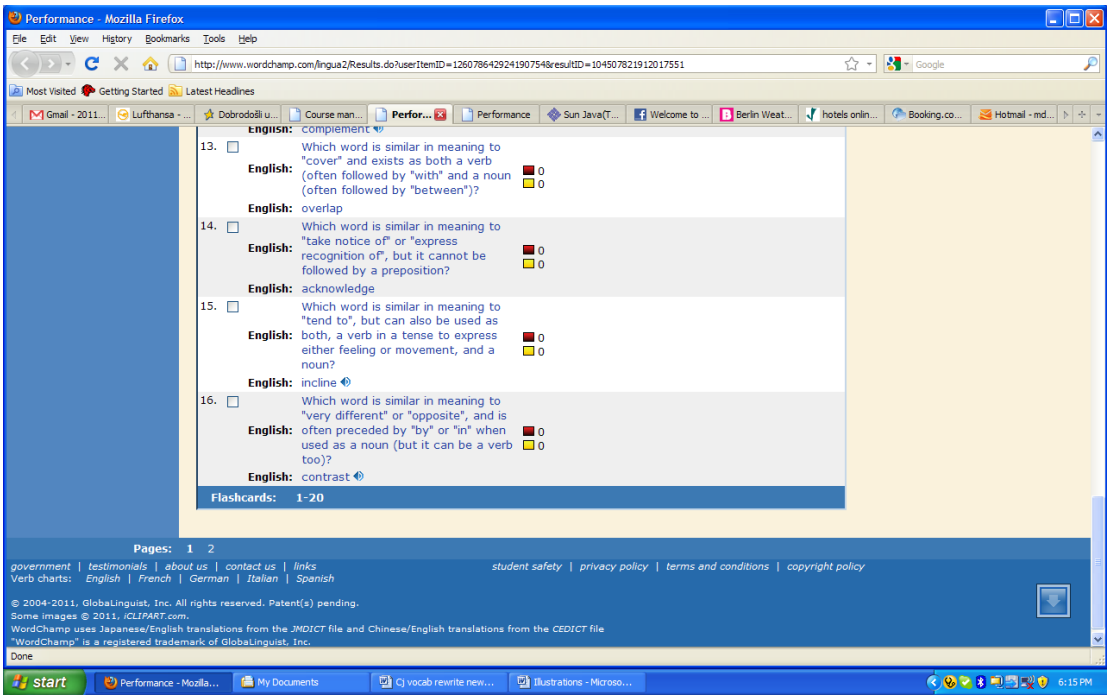

Figure 2: WordChamp activity log

Results

The raw data, not presented here, consist of several columns: 1) Versant Test of English, 2) initial VLT first 2000 words score, 3) initial VLT AWL score 4) results for the contrast procedure, i.e. vocabulary learning without flashcards, 5) results for the first treatment procedure, i.e. vocabulary learning with teacher-designed cards, 6) results for the second treatment procedure, i.e. vocabulary learning with student-designed cards, 7) final VLT AWL score. After sets with an incomplete vocabulary score sequence were removed, 89 data sets remained.

Table 1: ANOVA Data Summary

	Samples					
	1	2	3	4	5	Total
N	89	89	89			267
ΣX	5518	5474	4490.5			15482.5
Mean	62	61.5056	50.4551			57.9869
ΣX^2	384022	392184	291629.25			1067835.2
Variance	476.2045	630.7074	739.3275			639.2978
Std.Dev.	21.8221	25.1139	27.1906			25.2843
Std.Err.	2.3131	2.6621	2.8822			1.5474

Table 2: ANOVA Analysis Summary

standard weighted-means analysis					
ANOVA Summary Correlated Samples k=3					
Source	SS	df	MS	F	P
Treatment [between groups]	7584.1367	2	3792.0684	13.67	<.0001
Error	48828.6966	176	277.4358		
Ss/Bl	113640.3708	88			
Total	170053.2041	266			

An initial one-way analysis of variance (ANOVA) for repeated measures, performed using the VassarStats online software package (Lowry, 2012), indicated that the participants differed significantly in their performance under the three test conditions (columns 4, 5 & 6). The values for these columns are summarized in Table 1, whereas Table 2 contains an ANOVA analysis summary, with the key indicator for difference. This indicator, $F = 13.67$ ($p < .0001$), suggests that the learning condition effect was statistically significant.

Table 3: Tukey's HSD test and effect size

Comparison	Difference between means	Significant for HSD = 7.37, $p < 0.01$	Cohen's d
Conditions 1 and 2	0.4944	No	0.021
Conditions 1 and 3	11.5449	Yes	0.422545
Conditions 2 and 3	11.0505	Yes	0.471098

Tukey's HSD (honestly significant difference) test was subsequently used for multiple comparisons between groups of results (Table 3). While there was no significant difference between the contrast procedure and teacher-designed cards ($q = 0.18$, $p < 0.01$), student-designed cards yielded significantly lower results than either the contrast procedure ($q = 4.12$, $p < 0.01$) or the teacher-designed cards ($q = 4.31$, $p < 0.01$). A Bonferroni correction set the required alpha at 0.016, indicating that the student-designed cards had no effect, as opposed to teacher-designed cards and the contrast procedure.

To assess the practical significance of this finding, Cohen's f was first used to identify the magnitude of difference among groups. At $f = 0.55421$ Cohen's f was found to be large (Cohen, 1988). Next, Cohen's d was used to determine the effect size for the differences in pairwise comparisons (Table 3). According to Cohen's (1988) categories (0.2 small, 0.5 medium, and 0.8 large), the two statistically significant contrasts in this study seem to have small-to-medium, but not trivial, effect size (0.422, 0.471), while the statistically insignificant comparison has a trivial effect size (0.021).

Table 4: AWL comparison

AWL	Initial	Final
Mean	44.23	67.17
SD	19.64	21.04
SEM	2.42	2.59
N	66	66

Table 5: AWL values

Category	Value
t	9.2796
df	65
SED	2.472
p	< 0.0001

Overall, comparing the initial and final VLT AWL scores (Tables 4 & 5), a t-test established a statistically significant growth in academic vocabulary ($t = 9.2796$, $p < 0.0001$) over the course of the entire study, with averages from 44.23% to 67.17%, using the 66 complete result sets with both AWL test scores. However, mastery of AWL, set at 80% (Nation, 2006), remained yet to be achieved. While both the initial ($r = 0.70061701$, $p < 0.001$) and the final ($r = 0.50076793$, $p < 0.001$) AWL score correlated at a statistically significant level (Table 6) with the Versant score as a measure of English proficiency (Schacht & Aspelmeier, 2005), none of the post-test vocabulary results correlated significantly with the Versant scores (Table 6). This suggests that a factor other than general English proficiency was at work in the target vocabulary learning measures.

Table 6: Correlations

Versant relationships	Pearson r
V AWL1	0.70061701
V w1	0.11358256
V w2	0.28315694
V w3	0.36032762
V AWL2	0.50076793

Discussion

The students obtained the lowest scores when using self-designed cards, a procedure that was the least favorite in the classroom situation in this study. In light of some of the literature reviewed in a previous section of this document, this is a not entirely unexpected outcome, possibly caused by an affective barrier (Ortega, 2010; Krashen, 2004). At the same time, this outcome can be regarded as unexpected, as sources generally point to the potential learning benefits of creating one's own word cards (Nation, 2010; Schmitt, 2010). To explore every avenue, several other factors were vetted as possible causes, including the time spent practicing with cards, the number of overridden cards, the card design, cognitive load, nature of the task, skill transfer and fatigue.

The first line of enquiry was to search the WordChamp activity logs for possible evidence of omitted or reduced drill activity with student-designed cards. No such

evidence could be found. However, the information regarding cards answered incorrectly or overridden was not available for student-designed cards, while it was available for teacher-designed cards. This is one of the limitations of this study. Consequently, it is not entirely clear whether the manner in which the drills were carried out was responsible for the outcome.

While the information regarding the time spent designing the cards was not available, card design effectiveness was examined as a factor possibly influencing the results on the post-test. For this reason, a teacher looked at drills with ten random student-designed card stacks, and was only able to answer correctly less than two-thirds, while correctly answering almost all of the cards designed by another teacher. Upon closer examination, differences could be identified between teacher-designed and student-designed cards. While the teacher-designed cards integrated dictionary and concordance information using the lexis and notions familiar to the students, the student-designed cards, although including some semantic and grammatical information, contained hardly any contextual clues (see Figure 3). Folse (2006) points out that students tend to have problems with vocabulary in context, including collocations, use and meaning, and hence seem prone to making less informed use of such features in their word card designs. Moreover, the information presented in student-designed cards was less coherent, thus making these cards a less effective learning tool. This is based on the observation made through research that low-knowledge readers do not benefit from less coherent texts (Schnotz & Kurschner, 2007).

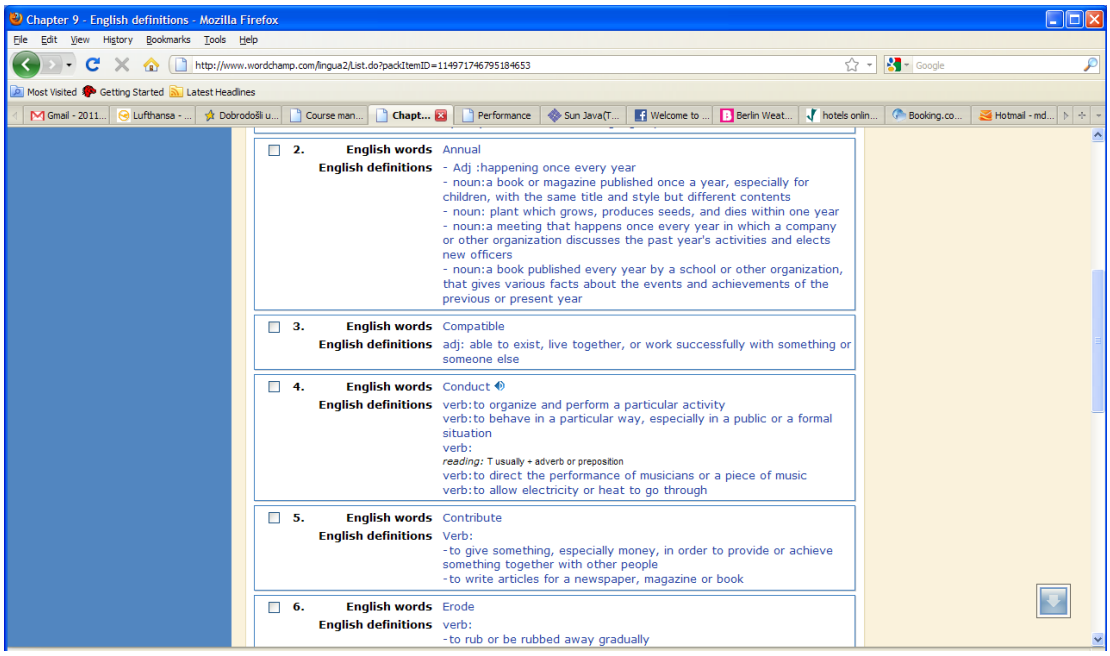

Figure 3: Student-created cards in the design mode

Furthermore, the activity of developing cards could have increased the extraneous cognitive load, i.e. the load on the working memory exerted by components other than the target vocabulary (Sweller, 2010). It is not clear whether the cards were less

effective because of the intrinsic cognitive load (Sweller, 2010) caused when processing semantic, lexical, grammatical and contextual word information at the same time, and trying to master simultaneously what is usually learned gradually (Schmitt, 2010) or because of the extraneous cognitive load (Sweller, 2010) caused by the use of dictionaries, concordances and electronic cards. Further qualitative research involving interviews with student card designers or think-aloud protocols during the card design process would be necessary to answer this question. One of the limitations of the present study is the absence of such qualitative research.

It is also not clear whether L2-L1 card design would have led to a different outcome, which remains another question for further research. The mere fact that learning with self-designed cards is a two-step process, consisting of design and review, seems to cause an increase in extraneous cognitive load, especially since the design part is a problem-solving activity. Problem-solving is associated with a higher extraneous cognitive load than worked examples (van Gogh et al., 2010), which translates as ready-made or teacher-designed cards in this case. It would also be interesting to see whether student-designed cards would have yielded a better outcome had they been collaboratively designed by groups of students, as collaborative learning is deemed to be more conducive to problem-solving (van Gogh & Rummel, 2010), and has already been successfully deployed in the creation of collaborative word banks (Horst, Cobb & Nicolae, 2005). This is another question for further research.

One way of determining whether the card design activity was too difficult is to examine the correlation between the available initial English proficiency scores and the learning outcome evidence from the card-supported activities (Schnotz & Kuerschner, 2007). However, the correlation coefficients between the Versant scores and the outcomes for each of the three conditions were similarly low and there was no statistical significance in any of the three coefficients (Table 6). Consequently, it seems that the level of English proficiency was not responsible for the outcome. However, future research might include a wider variety of parts of speech, in order to achieve a broader claim of generality for this conclusion.

Two possible issues that can affect the results in a within-sample design are the transfer of skills or learning from one task to the next, on the one hand, and fatigue on the other (Felix, 2008). In the current study, the two effects were, however, minimized by varying the order of both conditions and vocabulary groups. Thus, it seems less likely that the results could be attributed to either of the confounding variables, carryover or fatigue.

Furthermore, no information was collected regarding the cognitive styles of participants, leaving the door open to speculation that learning styles (Oxford & Crookall, 1990) might have been responsible for the results. This consequently raises another question for further research. One final limitation of this study is that no information was collected regarding the alternative vocabulary learning strategies used in the contrast procedure, making this one more project for future investigation.

Overall, within the confines of this study, it seems that teachers did create more effective electronic word cards than their students. More generally speaking, teachers tend to have an advantage in terms of skills, when it comes to design of learning material. A greater knowledge of teaching methodology as well as a better understanding of the vocabulary learning processes is one possible explanation for the success of their card design. A question to ask in future research might be what knowledge and skills are needed in order to create effective electronic word cards.

Implications for Pedagogy

Traditional flashcards are an important tool in vocabulary learning. Research indicates that decontextualizing techniques, such as L2-L1 cards or word-definition cards, when transferred to the computer, are equally as effective as traditional word cards (Nakata, 2008; Christensen et al., 2007). There is also some evidence that a semi-contextualizing strategy, i.e. pictorial information added to electronic mobile phone cards, can be conducive to vocabulary learning. However, the cards used in most studies have not been designed by students.

While the literature (Oxford & Crookall, 1990; Nation, 2006) encourages learner-designed conventional decontectualized word cards, no such advice is given regarding electronic word cards. In particular semi-contextualizing cards seem complex and therefore more difficult to design than decontextualizing word cards. In comparison with teacher-designed cards, student-designed electronic word cards seem to include less useful information and appear to be less coherent. Two approaches to this problem are possible: resorting to teacher-designed cards, on the one hand, and explicitly teaching the principles of effective card design to students, on the other. The choice would depend on the context for the instruction, including factors such as time available, level of proficiency and instructional goals. If, for example, the time allotted for English classes is limited and fast progress is expected, one might argue that this time would be better spent in learning vocabulary than in learning strategies for creating effective cards. If, on the other hand, time is available and one of the instructional goals is to equip the learners for future independent learning, then teaching them how to create effective word cards might be important. Students of lower English proficiency levels may need to abstain from creating semi-contextualizing electronic word cards and be encouraged to create contextualizing cards, including L2-L1 and word-picture or word-video cards.

Conclusion

The research question in this study has been: Do students benefit more from teacher-designed electronic flashcard activities than they do from those they designed themselves? A within-sample design with a contrast and two treatment procedures was employed to answer that question. The results indicate that the students learn best when using either their own favorite strategies or teacher-designed cards. Upon investigation, student-designed cards used semi-contextualizing information less skillfully than did the teacher-designed cards, leading to the conclusion that the semi-contextualizing features are complex and more training is needed for students to master these techniques than was originally anticipated. While designing one's own

cards may certainly lead to deeper processing, it is the number of word retrievals that leads to word retention (Folse, 2006). Therefore, it is the conclusion of this study that students may benefit more from teacher-designed semi-contextualizing electronic flashcards. To further investigate this claim, however, a study including a greater variety of parts of speech would be needed. It would also be of interest to perform a comparative study on decontextualized electronic word cards. In addition, further research into learning styles and the strategies used by students to design electronic cards, as well as into the skills required to design such cards would lead to a better understanding of the issues at hand.

References

Alderson, J. C. (2005). *Diagnosing foreign language proficiency*. London: Continuum.

Algina, J., Keselman, H. J., & Penfield, R. D. (2005). Effect Sizes and their Intervals: The Two-Level Repeated Measures Case. *Educational and Psychological Measurement*, 65, 241 – 258.

Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning and Technology*, 5 (1), 202 – 232.

Baker, R. L. (1984). Foreign language software: the state of the art or pick a card, any (flash) card. *CALICO Journal*, 2 (1), 6 – 27.

Bush, M. D., & Browne, J. M. (2004). Teaching Arabic with technology at BYU: learning from the past to bridge the future. *CALICO Journal*, 21 (3), 497 – 522.

Canado, M. L. P. (2010). Using virtual learning environments and computer-mediated communication to enhance the lexical competence of pre-service English teachers: a quantitative and qualitative study. *Computer Assisted Language Learning*, 23 (2), 129 – 150.

Chen, N. S., Hsie, S. W. & Kinshuk (2008). Effects of short-term memory and content representation type on mobile language learning. *Language Learning and Technology*. 12 (3), 93 – 113.

Christensen, E., Merril, P., & Yanchar, S. (2007). Second language vocabulary acquisition using a diglot reader or a computer-based drill and practice program. *Language Learning and Technology*, 20 (1), 67 – 77.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Coxhead, A. (2000). The academic word list. TESOL Quarterly. 34 (2), 213 – 238.

Ellis, N. (2001). Memory for language. In P. Robinson (ed.) *Cognition and Second Language Instruction* (pp. 33 – 68). Cambridge: Cambridge University Press.

Ellis, R. (1997). *SLA Research and Language Teaching*. Oxford: Oxford University Press.

Felix, U. (2008). The unreasonable effectiveness of CALL: what have we learned in two decades of research? *ReCALL*, 20 (2), 141 – 161.

Folse, K. S. (2004). *Vocabulary myths*. Ann Arbor: Michigan University Press.

Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly*, 40 (2), 273 – 293.

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.

Godwin-Jones, R., A. (2011). Emerging technologies: mobile apps for language learning. *Language Learning and Technology*. 14 (2), 2 – 11.

Horst, M., Cobb, T., & Nicolae, I. (2005). Expanding academic vocabulary with an interactive online database. *Language Learning and Technology*, 9 (2), 90 – 110.

Kaur, J., & Hegelheimer, V. (2005). ESL students' use of concordance in the transfer of academic word knowledge: An exploratory study. *Computer Assisted Language Learning*. 18 (4), 287 – 310.

Krashen, S. (2004). *The power of reading,* 2nd Edition. Littleton: Libraries Unlimited.

Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: do we need both to measure vocabulary knowledge? *Language Testing*, 21 (2), 202 – 226.

Lew, M. D. N., Alwis, W. A. M., & Schmidt, H. G. (2010). Accuracy of students' self-assessment and their beliefs about its utility. *Assessment & Evaluation in Higher Education*, 35 (2), 135 – 156.

Lowry, R. (2012). VassarStats: Website for statistical computation. Retrieved on 10 February 2012 from http://faculty.vassar.edu/lowry/VassarStats.html

McDonough, J., & McDonough, S. (1997). *Research methods for English language teachers*. London: Arnold.

Mohsen, M. A., & Baukumar, M. (2011). A review of multimedia glosses and their effects on L2 vocabulary acquisition in CALL literature. *ReCALL*. 23 (2), 135 – 159.

Nakata, T. (2008). English vocabulary learning with word lists, word cards and computers: implications for cognitive psychology research for optimal spaced learnig. *ReCALL*. 20 (1), 3 – 20.

Nassaji, H. (2003). L2 vocabulary learning from context: strategies, knowledge, sources and their relationship with success in L2 lexical inferencing. *TESOL Quarterly*. 37 (4), 645 – 670.

Nation, I. S. P. (1990). *Teaching & learning vocabulary*. Boston: Heinle.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P. (2006). Vocabulary: Second language. In K. Brown (ed.) *Encyclopaedia of Language and Linguisti*cs, 2nd Ed. Oxford: Elsevier. Vol 13: 448-454.

Nation, I. S. P. (2008). *Teaching vocabulary: Strategies and techniques*. Boston: Heinle.

Ortega, L. (2010). *Second language acquisition: Critical concepts in linguistics*. London: Routledge.

Oxford, R., & Crookall, D. (1990). Vocabulary learning: A critical analysis of techniques. *TESL Canada Journal*, 7 (2), 9 – 30.

Paribakht, T. S., & Wesche, M. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal*, 11 (1), 9 – 29.

Svalberg, A.M-L. (2007). Language awareness and language learning. *Language Teaching*, 40, 287 – 208.

Sawyer, M., Ranta, L. (2001). Aptitude, individual differences, and instructional design. In P. Robinson (ed.) *Cognition and second language instruction* (pp. 319 – 353). Cambridge: Cambridge University Press.

Schacht, S. P., & Aspelmeier, J. E. (2005). *Social and behavioral statistics: A user-friendly approach*. Boulder: Westview Press.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research,* 12 (3), 329 – 363.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual.* London: Palgrave Macmillan.

Schmitt, N., Ng, J. W. C., & Garrad, J. (2011). Word associates format: Validation evidence. *Language Testing,* 12 (3), 329 – 363.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two new versions of the Vocabulary Levels Test. *Language Testing,* 28 (1), 105 – 126.

Schnotz, W., & Kurschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review.* 19, 469 – 508.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germaine cognitive load. *Educational Psychology Review.* 22, 123 – 138.

Van Gog, T., Paas, F & Sweller, J. (2010). Cognitive load theory: Advances in research on worked examples, animations and cognitive load measurement. *Educational Psychology Review.* 20, 375 – 378.

Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review.* 22, 155 – 174.

Won, M. (2008). The effects of vocabulary instruction on English language learners: meta-analysis. Doctoral Dissertation, Texas Tech University.

Appendix

Modified Vocabulary Knowledge Scale (VKS)

1. I know/ don't know this word. It means _____.
   (If you know the word, underline "know" above and provide an English synonym or a translation in your native language. Otherwise, underline "don't know".)
2. I can use this word in a good example sentence. Provide your sentence here:
   _____
   (If you do #2, you must also provide a synonym or translation for #1.)