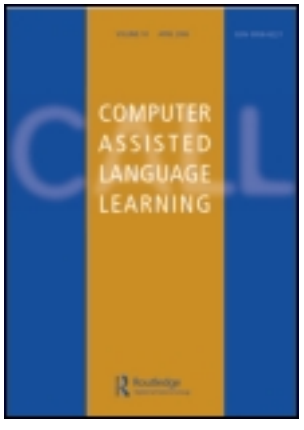
This article was downloaded by: [Victoria University of Wellington]
On: 25 September 2012, At: 21:51
Publisher: Routledge
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Computer Assisted Language Learning
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ncal20

Computer-assisted second language vocabulary learning in a paired-associate paradigm: a critical investigation of flashcard software

Tatsuya Nakata [a]

[a] School of Linguistics and Applied Language Studies, Victoria University of Wellington, Wellington, New Zealand

Version of record first published: 25 Jan 2011.

To cite this article: Tatsuya Nakata (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

To link to this article: http://dx.doi.org/10.1080/09588221.2010.520675

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Computer Assisted Language Learning
Vol. 24, No. 1, February 2011, 17–38

Routledge
Taylor & Francis Group

Computer-assisted second language vocabulary learning in a paired-associate paradigm: a critical investigation of flashcard software

Tatsuya Nakata*

School of Linguistics and Applied Language Studies, Victoria University of Wellington, Wellington, New Zealand

Downloaded by [Victoria University of Wellington] at 21:51 25 September 2012

The present study aims to conduct a comprehensive investigation of flashcard software for learning vocabulary in a second language. Nine flashcard programs were analysed using 17 criteria derived from previous studies on flashcard learning as well as paired-associate learning. Results suggest that in general, most programs have been developed in a way that maximises vocabulary learning. For instance, seven of the nine programs allow flashcard creation, offer multilingual support and allow learners to add contexts, audios or images to flashcards. Furthermore, eight programs provide various types of exercises, and nine programs support scheduling. At the same time, the present study has also shown that existing flashcard programs have some room for improvement. More specifically, most programs are limited in their ability to support data entry, increase retrieval effort and promote generative use of target words. When individual programs are compared, *iKnow*! seems to be the best program among those investigated. It offers the most comprehensive support for data entry, automatically generates distractors for multiple-choice exercises and increases retrieval effort by systematically introducing various types of exercises. The variations among the programs in their design suggest that there do not exist commonly accepted guidelines for how flashcard software should be designed.

Keywords: flashcard learning; vocabulary; CALL; paired-associate learning; retrieval

Introduction

The present study conducts a comprehensive investigation of flashcard software for learning vocabulary in a second or foreign language (L2). Flashcard programs are defined as software that encourages learners to study L2 vocabulary in a *paired-associate* format. In other words, in these programs, target items are presented outside meaning-focused tasks, and learners are asked to associate the L2 word form with its meaning, usually in the form of a first language (L1) translation, L2 synonym, or L2 definition.

The present study is motivated by several pedagogical and practical concerns. First, although paired-associate learning, including learning from flashcards, tends to be dismissed as a relic of the old-fashioned behaviourist learning model (Hulstijn,

*Email: tatsuya.nakata@vuw.ac.nz

ISSN 0958-8221 print/ISSN 1744-3210 online
© 2011 Taylor & Francis
DOI: 10.1080/09588221.2010.520675
http://www.informaworld.com

Downloaded by [Victoria University of Wellington] at 21:51 25 September 2012

2001), empirical studies demonstrate that it is effective and efficient. Studies have shown that in a paired-associate learning task, large numbers of words can be memorised in a very short time (e.g. Fitzpatrick, Al-Qarni, & Meara, 2008; Nation, 1980; Thorndike, 1908). Vocabulary learnt in a paired-associate format is also resistant to decay (Fitzpatrick et al., 2008; Thorndike, 1908) and can be retained over several years (H.P. Bahrick, L.E. Bahrick, A.S. Bahrick, & P.E. Bahrick, 1993; H.P. Bahrick & Phelps, 1987). Recent studies have also suggested that flashcard learning may transfer to normal language use and is a valuable learning activity (Elgort, 2007; Webb, 2002, 2009a). Given the effectiveness and efficiency of flashcard learning, it is useful to conduct a comprehensive investigation of existing flashcard programs to examine whether they have been developed in a way that maximises vocabulary learning.

Second, some researchers argue that computer-based flashcards may allow learners to learn more effectively than paper-based ones because the former offer benefits that the latter do not. For instance, computers can be easily programmed to keep track of a learner's performance and control the sequencing of items to make sure that unknown or hard items are studied more frequently than known or easy items (e.g. van Bussel, 1994; Nakata, 2008; Pyc & Rawson, 2007, 2009; Siegel & Misselt, 1984), which can be cumbersome if done manually. Other advantages offered by computer-assisted language learning (CALL) may include enhanced presentation of materials due to its multimedia capabilities, introduction of new exercise types and positive effects on students' motivation (e.g. Allum, 2004; Ellis, 1995; Garcia & Arias, 2000; Hulstijn, 2001; Nation, 2001, pp. 108–110; Nesselhauf & Tschichold, 2002). A comprehensive survey of flashcard software needs to be conducted to determine whether existing programs have exploited advantages of computer-based instruction to the fullest.

Third, an informal survey of computer-based flashcards reveals that there are discrepancies in the way they are designed. For instance, programs differ in the editing flexibility, types of exercises, or review schedule. The inconsistency in the design of the software suggests that there do not exist commonly accepted guidelines for developing effective flashcard programs. By conducting a comprehensive investigation of flashcard software, this study may help establish guidelines for how flashcard programs should be designed.

Lastly, some computer-based flashcards have been used very widely. For instance, *vTrain*, a flashcard program, has been used by more than 50 universities and hundreds of schools worldwide (Rädle, 2009) while *Quizlet* has more than one million registered users (Quizlet, 2010). In Yawata City in Kyoto, Japan, all the public junior high schools have incorporated into their English curriculum a flashcard program for *Nintendo DS*, a portable game player (Tamaki, 2007). Given the widespread use of computer-based flashcards, it seems useful to examine the pedagogical value of these programs.

Even though the above concerns call for a comprehensive survey of flashcard software, there does not seem to have been any attempt to conduct a systematic analysis of flashcard programs. Although there have been some extensive investigations of computer-based dictionaries (Nesi, 1996; Rizo-Rodríguez, 2008), concordancer programs (Higgins, 1991), or vocabulary-building software (Nesselhauf & Tschichold, 2002), none of them has looked at flashcard programs. Furthermore, existing surveys on computer-based flashcards have examined only one program (e.g. Böhlke, 2002; Burston, 2007; Foster, 2009; Hsieh & Fei, 2009;

Ishikawa, 2004; Olmanson, 2007; Walker, 2006) and are not comprehensive. While these studies are useful, there is greater value in contrasting a wide range of programs to determine which software may provide the most benefit to users.

Literature review

In order to identify design features that contribute to ideal flashcard software, this section reviews previous studies on flashcard learning as well as paired-associate learning. Design features related to (a) flashcard creation and editing and (b) learning will be discussed. This review of literature will form a basis for criteria for evaluating flashcard software in the present study (see Criteria for evaluating the software).

Flashcard creation and editing

Regarding flashcard creation and editing, previous studies suggest that positive features of flashcard programs include a flashcard creation function, multilingual support, support for multi-word units, the ability to handle various types of information such as contexts, images, or audios, support for data entry, and support for flashcard sets (e.g. Böhlke, 2002; Burston, 2007; Foster, 2009; Hsieh & Fei, 2009; Ishikawa, 2004; Olmanson, 2007; Walker, 2006). Literature supporting each feature will be reviewed below.

Flashcard creation

An ideal flashcard program would allow learners to create their own flashcards (Böhlke, 2002; Burston, 2007; Foster, 2009; Hsieh & Fei, 2009; Walker, 2006). Some flashcard programs come with a wide selection of readymade flashcards for various languages (Böhlke, 2002; Burston, 2007; Foster, 2009; Hsieh & Fei, 2009; Olmanson, 2007; Walker, 2006). By using these flashcards, learners can study many vocabulary items while avoiding the time-consuming task of flashcard creation. Good flashcard software, however, should not only provide readymade flashcards but also offer a flashcard creation function. Such a function would be useful for learners who want to study technical or low frequency vocabulary for which readymade flashcards are often not available.

Multilingual support

It is desirable that a flashcard program allow learners to create both target items and their translations in any language (Böhlke, 2002; Burston, 2007; Walker, 2006). Multilingual support is beneficial for three reasons. First, it will allow learners to study various non-alphabet-based languages such as Japanese, Chinese, Arabic, or Thai (Böhlke, 2002; Burston, 2007; Hsieh & Fei, 2009; Walker, 2006). Second, previous research shows that use of L1 translations facilitates vocabulary learning (Lado, Baldwin, & Lobo, 1967; Laufer & Shmueli, 1997; Mishima, 1967). Multilingual support, therefore, should contribute to improved performance because it will enable learners of any first language background to use L1 translations. Third, L1 translations offer a more favourable condition for flashcard learning because low level learners may not be able to understand L2 definitions or synonyms.

Downloaded by [Victoria University of Wellington] at 21:51 25 September 2012

Multi-word units

The recent developments in corpus linguistics have shown that L2 learners have to acquire a large number of multi-word units to be able to produce and comprehend ideas accurately and fluently (Pawley & Syder, 1983; Wray, 2000, 2002). An ideal program, therefore, would let learners create flashcards for multi-word units as well as single words (Böhlke, 2002; Burston, 2007; Foster, 2009; Olmanson, 2007). Steinel, Hulstijn, and Steinel (2007) found that multi-word units such as idioms can be learnt effectively in a paired-associate format.

Types of information

Since vocabulary acquisition involves much more than associating new L2 words with their meaning (e.g. Beheydt, 1987; Nation, 1990, pp. 29–50, 2001, pp. 23–59), it is desirable that information such as collocations, contexts, or pronunciation (in the form of audio files) can be added to flashcards besides the word meanings. Software should also enable learners to add images or videos to flashcards because visual information facilitates vocabulary learning (e.g. Chun & Plass, 1996, 1997; Lado et al., 1967; Webber, 1978).

Support for data entry

Creating flashcards requires a considerable amount of time and energy on the part of learners. One advantage of computer-based flashcards over paper-based ones is that computers can help flashcard creation by automatically supplying information about lexical items (Burston, 2007; Foster, 2009; Ishikawa, 2004). Ideal software would allow learners to import meaning, parts of speech, contexts, or audio recordings of target words from an internal database or external resources. It would also be valuable if software could automatically supply frequency information derived from a corpus such as the British National Corpus or give a list of words in the same word family (e.g. *navigation, navigational, navigator* and *circumnavigate* for *navigate*). Frequency information will be useful because it gives learners a good indication of how useful a word is (Nation, 2001, pp. 6–22, 2008, pp. 7–15). Learning words from the same word family may help learners to efficiently increase their vocabulary size (Sökmen, 1992).

Flashcard set

Researchers argue that learning differences between semantically close words lead to a more precise understanding of each individual item (e.g. Beheydt, 1987; Nation, 2001, p. 103; Stahl & Nagy, 2006, pp. 77–96). For instance, in order to fully grasp what *amaze* means, one needs to be able to distinguish it from its synonyms such as *surprise, astonish, astound*, or *flabbergast*. Learning semantically related words simultaneously while focusing on their differences, therefore, is expected to contribute to a deep understanding of word meanings. With this in mind, ideal flashcard software would allow learners to create their own sets of flashcards (e.g. items related to numbers, colours, animals, or food) so that learners could review words belonging to the same semantic category. One caveat to be considered, though, is that learning semantically related words simultaneously inhibits learning

Downloaded by [Victoria University of Wellington] at 21:51 25 September 2012

of unfamiliar vocabulary because it often causes interference between words (e.g. Tinkham, 1993, 1997; Waring, 1997a). Therefore, it is advisable to study words in semantic sets only when most items being studied are already familiar to learners (Nation, 2001, p. 103; Stahl & Nagy, 2006, pp. 92–93).

Learning

Regarding design features related to learning, research suggests that ideal flashcard software would have presentation and retrieval modes, provide various types of exercises, increase retrieval effort, promote generative use, be flexible about the block size and support scheduling (e.g. Barcroft, 2002, 2004; Bjork, 1994, 1999; Kornell, 2009; Nation, 1982, 2001, pp. 68–70, 2008, p. 109; Pyc & Rawson, 2007, 2009). The rationale behind each design feature will be given below.

Presentation and retrieval modes

Retrieval practice, where learners are required to recall or recognise the L2 word form or its meaning, is found to yield superior retention than mere presentation (e.g. Barcroft, 2007; McNamara & Healy, 1995; Royer, 1973) because it strengthens retrieval routes to memory (Baddeley, 1997, p. 112; Ellis, 1995; Nation, 2001, p. 79). At the same time, it has been shown that only a successful retrieval strengthens memory (Modigliani, 1976). These findings imply that learners should be introduced to unfamiliar target words first and then tested on their knowledge of these partially known words. An ideal flashcard program, therefore, would consist of two modes: a presentation mode, where learners familiarise themselves with the target words, and retrieval mode, where they practise retrieval of previously met words (Nation, 1982). A presentation mode would also be valuable because asking learners to practise retrieval of totally unfamiliar words would result in unsuccessful performance and have negative effects on their motivation.

Receptive recall, receptive recognition, productive recall and productive recognition

Retrieval practice can be categorised into four types: receptive recall, receptive recognition, productive recall and productive recognition (Laufer, Elder, Hill, & Congton, 2004; Laufer & Goldstein, 2004). In receptive recall, learners are asked to produce the meaning of target words while in productive recall, they produce the target word form corresponding to the meaning provided. Receptive recognition requires learners to choose, rather than to produce, the meaning of target words, whereas productive recognition requires learners to choose the target word form corresponding to the meaning provided (Laufer et al., 2004; Laufer & Goldstein, 2004). Previous studies suggest that a good flashcard program should support more than one type of retrieval practice for at least two reasons. First, it has been demonstrated that receptive retrieval promotes larger gains in receptive knowledge while productive retrieval is effective for gaining productive knowledge (e.g. Griffin & Harley, 1996; Waring, 1997b; Webb, 2002, 2009b). Therefore, in order to gain both receptive and productive vocabulary knowledge efficiently, learners need to practise receptive as well as productive retrieval. Second, according to the type of processing-resource allocation (TOPRA) model (Barcroft, 2002, 2004), it is difficult for learners to acquire both the word form–meaning connection and the word form

Downloaded by [Victoria University of Wellington] at 21:51 25 September 2012

(namely, spelling and pronunciation) of a word simultaneously due to their limited cognitive resources. The TOPRA model implies that flashcard software should provide at least two types of exercises: one that focuses on linking meaning with form such as receptive recognition, receptive recall and productive recognition and the other that directs learners' attention to the word form, that is, productive recall.

Increasing retrieval effort

According to the *retrieval effort hypothesis*, the degree to which a successful retrieval enhances memory increases with the difficulty of the retrieval practice (Bjork, 1994, 1999; Pyc & Rawson, 2009). This hypothesis implies that it is desirable for a flashcard program to arrange various types of exercises in the order of increasing difficulty. In other words, ideal software would test learners' vocabulary knowledge in a relatively easy format such as receptive recognition or productive recognition in earlier stages and introduce a more demanding one such as receptive recall or productive recall later.

Generative use

In order to increase vocabulary learning, it is also essential for flashcard software to encourage *generative use* of words, where learners encounter or use previously met words in novel contexts (Joe, 1995, 1998; Nation, 2001, pp. 68–70). For instance, let us suppose that learners first encountered the word *break* in the sentence 'He broke the cup.' Learners may hypothesise that *break* can take only concrete nouns as the object. Exposure to expressions such as *break the record, break a promise*, or *break the news* may help learners to reconceptualise and further deepen their knowledge of the word. Computers can facilitate generative use of words effectively because they can present materials in an organised fashion so that various aspects of knowledge of particular words are systematically introduced to learners (Groot, 2000). A good flashcard program, therefore, would show the target word used in different senses, collocations, inflections, grammatical functions, or sentence patterns every time the word is practised.

Block size

The block size is defined as the number of cards to be studied in one learning session. There exist several conflicting views about what constitutes the optimal block size. The *spacing effect* predicts that a large block size is more effective than a small one. According to this effect, the larger the intervals between study opportunities for a given item, the better the retention will be (e.g. Baddeley, 1997, pp. 108–114; H. P. Bahrick et al., 1993; H. P. Bahrick & Phelps, 1987; Kornell, 2009). The spacing effect favours a larger block size because it increases the intervals between learning opportunities (Kornell, 2009). For instance, when the block size is five, only four items are encountered between study trials of a given item. In contrast, when the block size is 100, 99 items are encountered between study trials of a given item, contributing to longer intervals between repetitions and consequently, better learning (Kornell, 2009).

In contrast, the *retrieval practice effect* and *list-length effect* suggest that a small block size is more effective. The retrieval practice effect refers to the phenomenon

Downloaded by [Victoria University of Wellington] at 21:51 25 September 2012

where a successful recall from memory yields superior retention than mere presentation of the target item (e.g. Barcroft, 2007; McNamara & Healy, 1995; Royer, 1973). A small block size is more likely to lead to retrieval success than a large one because in the former, most items will be tested before forgetting occurs. Therefore, the retrieval practice effect predicts that a small block size is more effective. The list-length effect, which states that memory performance is inversely related to the number of items in a list (Gillund & Shiffrin, 1984), also favours a small block size.

Empirical studies on the block size have yielded inconsistent results (e.g. van Bussel, 1994; Kornell, 2009; Pyc & Rawson, 2007, 2009), and researchers also disagree over the optimal block size. While Salisbury and Klein (1988) and van Bussel (1994) recommend a small block size, Kornell (2009) advocates using large stacks of flashcards. Nation (2008, p. 109) suggests a block size of 20 to 50, depending on the proficiency of the learners. Given the lack of consensus among researchers regarding the optimal block size, flashcard software should be flexible about the block size.

Downloaded by [Victoria University of Wellington] at 21:51 25 September 2012

Adaptive sequencing

Most new words will eventually be forgotten after only a single encounter, and in order for words to be remembered over time, they need to be reviewed on a regular basis (Baddeley, 1997, pp. 108–114; Ellis, 1995; Hulstijn, 2001). Computers can facilitate systematic review of lexical items with an *adaptive sequencing procedure*, or an algorithm to change sequencing of items based on learners' previous performance on individual items (e.g. van Bussel, 1994; Nakata, 2008; Pyc & Rawson, 2007, 2009; Siegel & Misselt, 1984). For instance, computers can keep track of a learner's performance and make sure that unknown or hard items are studied more frequently than known or easy items.

Expanded rehearsal

Among many kinds of adaptive sequencing procedures, *expanded rehearsal* is widely believed to be the most effective (e.g. Baddeley, 1997, pp. 112–114; Ellis, 1995; Hulstijn, 2001; Pimsleur, 1967). Expanded rehearsal refers to a review schedule where the intervals between study trials are gradually increased as learning proceeds. For instance, in expanded rehearsal, the first review takes place one day after the initial encounter, the second review a week after the first review, the third review two weeks after the second, and so forth.

Methodology

The review of literature has revealed a range of useful principles that can be applied when developing computerised vocabulary learning programs. Let us now see how well existing programs apply these.

Criteria for inclusion in the analysis

Currently, numerous flashcard programs are available both commercially and freely. Due to the large number of programs available, it is not very practical to investigate all the existing computer-based flashcards. Hence, programs to be included in the

Downloaded by [Victoria University of Wellington] at 21:51 25 September 2012

Table 1. Software to be evaluated in the present study.

	SuperMemo 2008	*vTrain 5.2*	*MemoryLifter 2.3*	*P-Study System Ver.8.3*	*WordChamp*	*Quizlet*	*iKnow!*	*Word Engine*	*LearnThatWord Vocabulary & Spelling Program Online*
Publisher	SuperMemo World	Paul Rädle	LearnLift	Hiroaki Takeuchi	GlobaLinguist, Inc.	Quizlet LLC	Cerego Japan Inc.	Lexxica	eSpindle Learning
URL	http://www.supermemo.com/	http://www.vtrain.net/	http://www.memorylifter.com/	http://www.takke.jp/	http://www.wordchamp.com/	http://quizlet.com/	http://smart.fm/	http://www.wordengine.jp/	http://www.LearnThatWord.org/
Price	US$50.00	US$25.00	Free	Free	Free, with optional US$9.95/year premium membership	Free, with optional US$10.00/year premium membership	Free	980 to 3980 yen per course	Pay-per-Result revenue system (3–50 US cents per successfully learnt item). Alternatively, US $79.80/year (auto-renewing), $99.80/year (non-renewing), $599 for lifetime membership
Target languages	Any language	Any language	Any language	Any language	Any language	Any language	Any language	English	English
Source languages	Any language	Any language	Any language	Any language	Any language	Any language	Any language	English and Japanese	38 languages (e.g. English, French, German, Spanish, Chinese)
Readymade flashcards	Available for over 25 languages (e.g. English, French, German, Spanish, Chinese); http://www.supermemory.com/sml/sml.htm	Available for 40 languages (e.g. English, French, German, Spanish, Chinese); http://www.vtrain.net/dbase.htm	Available for 18 languages (e.g. English, French, German, Spanish, Chinese); http://www.memorylifter.com/products.html	Available for over 13,000 English lexical items; http://www.takke.jp/pss/additional_questions.php	Available for over 140 languages; http://www.wordchamp.com/lingua2/Browse.do	Available for over 40 languages (e.g. English, French, German, Spanish, Chinese); http://quizlet.com/languages-and-vocab/	Available for 190 languages; http://smart.fm/explore	Available for around 78,000 English lexical items; http://www.wordengine.jp/courses	Available for 150,000 English lexical items; http://www.learnthat.org/about_us.html

(*continued*)

Downloaded by [Victoria University of Wellington] at 21:51 25 September 2012

Table 1. (*Continued*).

	SuperMemo 2008	*vTrain 5.2*	*MemoryLifter 2.3*	*P-Study System Ver.8.3*	*WordChamp*	*Quizlet*	*iKnow!*	*Word Engine*	*LearnThatWord Vocabulary & Spelling Program Online*
Supporting documentation	Available in English, Polish, Russian, Italian, Dutch, Czech, German, Serbian, and Portuguese	Available in English, French, German, and Spanish	Available in English, German, French, Spanish, and Portuguese	Available in Japanese	Available in English, French, Spanish, Chinese, and Arabic	Available in English	Available in English and Japanese	Available in English and Japanese	Available in English
Guidelines for using the program	See http://www.supermemo.com/english/smintro.htm and http://www.supermemo.com/help/guide.htm	See Böhlke (2002) and Walker (2006)	See Hsieh & Fei (2009)	See Ishikawa (2004)	See Burston (2007)	See Foster (2009)	See http://smart.fm/tour	See http://www.wordengine.jp/vflash and http://www.wordengine.jp/research/presentation	See Olmanson (2007)
The rationale behind the choice	Developed under the supervision of researchers	Reviewed favourably by Böhlke (2002) and Walker (2006)	Reviewed favourably by Hsieh and Fei (2009)	Reviewed favourably by Ishikawa (2004). The most downloaded flashcard program in Vector as of 3 May 2010.[a]	Reviewed favourably by Burston (2007)	Reviewed favourably by Foster (2009). Has more than a million registered users (Quizlet, 2010)	Developed under the supervision of researchers	Developed under the supervision of researchers	Reviewed favourably by Olmanson (2007)

Note: [a]Vector (http://www.vector.co.jp/) is a Japanese equivalent of download.com (http://download.cnet.com/) and has more than 110,000 freeware and shareware programs available for download as of 3 May 2010.

analysis have been selected based on several criteria. First, the software needs to meet all of the following four criteria:

(1) The program is flashcard-based. In other words, target items are presented outside meaning-focused tasks, and users study L2 vocabulary in a paired-associate format.
(2) The program is commercially or freely available. Software that is developed for research purposes and not made public was excluded.
(3) The program is developed for Windows or Macintosh computers. Programs developed for portable devices such as mp3 players or smartphones were excluded because although these devices are becoming increasingly popular, they are not as common as personal computers yet. Furthermore, compared with programs for PCs, those for mobile devices are limited in their capabilities.
(4) The program has been developed or updated after 2005. Older programs were excluded because some of them are not compatible with current operating systems such as Windows Vista or Windows 7. User support is often not available for some old programs either.

Additionally, the programs need to meet at least one of the following three criteria:

(1) The program has been developed under the supervision of researchers such as applied linguists, cognitive psychologists, or brain scientists. These programs were included in this analysis because research-based programs are assumed to be superior to non research-based ones.
(2) The program has been reviewed favourably in an academic publication.
(3) The program is popular. Popularity is measured in terms of the number of downloads or users (see Table 1).

Downloaded by [Victoria University of Wellington] at 21:51 25 September 2012

Software to be evaluated

Based on the above criteria, nine programs were identified. Table 1 summarises the names, publishers, URLs, prices, target languages, source languages, readymade flashcard availability, supporting documentation and guidelines for using the programs as well as the rationale behind the choice. Note that the last five programs (from *WordChamp* to *LearnThatWord*) are web-based and require an Internet connection. The author of this article has neither affiliation with nor financial interest in any of the nine programs.

Before presenting evaluation of the nine programs, it may be useful to clarify the target users of each program. All the programs except *Word Engine, LearnThatWord* and *P-Study System* are appropriate for learners of any language irrespective of their L1 because as we will see in the Results section, these six programs allow users to create their own flashcards in any language. The six programs also provide a wide selection of readymade flashcards for various languages including English, French, German, Spanish, Chinese, Russian and Arabic (Table 1).

P-Study System has been developed primarily for Japanese learners of English. Just like the six programs above, *P-Study System* offers a flashcard creation function and multilingual support and can be used for learning any language. Yet, the

Downloaded by [Victoria University of Wellington] at 21:51 25 September 2012

Table 2. Comparison of the nine flashcard programs.

	SuperMemo 2008	*vTrain 5.2*	*Memory Lifter 2.3*	*P-Study System Ver.8.3*	*WordChamp*	*Quizlet*	*iKnow!*	*Word Engine*	*Learn ThatWord*
Flashcard creation	+	+	+	+	+	+	+	No	No
Multilingual support	+	+	+	+	+	+	+	N/A	N/A
Multi-word units	+	+	+	+	+	+	+	N/A	N/A
Types of information	Meaning, context, audio, image, video, HTML, PDF, OLE object, etc.	Meaning, context, audio, image, video, OLE object, etc.	Meaning, context, audio, image, video	Meaning, audio, hint, phonetic symbol (e.g, IPA, pinyin, furigana), notes	Meaning, context, audio, image, phonetic symbol (e.g, IPA, pinyin, furigana)	Meaning, image	Meaning, parts of speech, context, audio, image, notes	N/A	N/A
Support for data entry	No	No	No	+ (Linked to web dictionaries)	+ (Meaning, audio, image)	+ (Meaning, image)	+ (Meaning, POS, context, audio, image)	N/A	N/A
Flashcard set	+	+	+	+	+	+	+	No	+
Presentation mode	+	+	+	+	+	+	+	No	No
Retrieval mode	+	+	+	+	+	+	+	+	+
Receptive recall	+	+	+	+	+	+	No	No	No
Receptive recognition[a]	+	+	+ +	+ +	+	+ +	+ +	+ +	No
Productive recall	+	+	+	+	+	+	+	No	+
Productive recognition[a]	+	+	+ +	+ +	+	+ +	+ +	+ +	No
Increasing retrieval effort	No	No	No	No	+	No	+	No	No
Generative use	No	No	No	No	No	No	No	No	No
Block size	Can be determined by the learner	Can be determined by the learner	Can be determined by the learner	Can be determined by the learner	10, 20, 30, 40, 50, 60, 70, 80, 90, 100 or unlimited (can be chosen by the learner)	Can be determined by the learner	5 or 10 (can be chosen by the learner)	10	15, 25, 50, 75, 100, 250, or 500 (can be chosen by the learner)
Adaptive sequencing	+	+	+	+	+	+	+	+	+
Expanded rehearsal	+	+	+	+	+	No	+	+	No

Note: [a] + + = The software supports multiple-choice exercises. Furthermore, (1) distractors are automatically generated by the program and (2) the software changes the position of the correct answer every time the word is tested; + = the software supports multiple-choice exercises. However, (1) distractors are not automatically generated by the program and (2) the software presents multiple choice options in a fixed order for a given item.

supporting documentation for *P-Study System* is available only in Japanese, and the software may not be suitable for non-Japanese readers. Furthermore, the program provides readymade flashcards only for English–Japanese translation pairs.

Word Engine and *LearnThatWord* are targeted solely at English learners. These two programs do not support flashcard creation, and their readymade flashcards are limited to English vocabulary. Consequently, users can study only English vocabulary with the two programs. *Word Engine* is primarily aimed at Japanese learners of English and can present the meanings of target items either in English or Japanese. *LearnThatWord* supports more source languages than *Word Engine*. In *LearnThatWord*, the meanings of target items can be given in 38 languages including English, French, German, Spanish, Chinese and Japanese.

As shown in Table 1, all the nine programs come with a wide selection of readymade flashcards. The readymade flashcards for all the programs except *Word Engine* and *LearnThatWord* have been created by developers of the software, users, or third party companies, and the selection criteria of items vary. The readymade flashcards for *Word Engine* and *LearnThatWord*, in contrast, have been prepared exclusively by the publishers of the software. *Word Engine* offers around 78,000 readymade flashcards for English lexical items (Lexxica, 2010). The items have been chosen based on frequency analysis of corpora assembled for specific subjects such as business English, Internet English, TOEFL preparation, Japanese university entrance examinations and so forth (J. Glick, personal communication, 28 April 2010). *LearnThatWord* comes with an internal database of 150,000 English lexical items, which have been selected based on several factors including frequency, difficulty and usefulness (R. Warda, personal communication, 2 August 2010).

Criteria for evaluating the software

Although some criteria for evaluating CALL software have been proposed (e.g. Chapelle, 1998; Plass, 1998; Rizo-Rodríguez, 2008), they are not appropriate for the purpose of the present study because they have not been developed for flashcard programs per se. Hence, the following checklist has been devised based on the design features that contribute to ideal flashcard software discussed in the Literature review.

Flashcard creation and editing

(1) Flashcard creation: Can learners create their own flashcards?
(2) Multilingual support: Can the target words and their translations be created in any language?
(3) Multi-word units: Can flashcards be created for multi-word units as well as single words?
(4) Types of information: Can various kinds of information be added to flashcards besides the word meanings (e.g. parts of speech, contexts, or audios)?
(5) Support for data entry: Does the software support data entry by automatically supplying information about lexical items such as meaning, parts of speech, contexts, or frequency information from an internal database or external resources?
(6) Flashcard set: Does the software allow learners to create their own sets of flashcards?

Downloaded by [Victoria University of Wellington] at 21:51 25 September 2012

Learning

(1) Presentation mode: Does the software have a presentation mode, where new items are introduced and learners familiarise themselves with them?
(2) Retrieval mode: Does the software have a retrieval mode, which asks learners to recall or choose the L2 word form or its meaning?
(3) Receptive recall: Does the software ask learners to produce the meanings of target words?
(4) Receptive recognition: Does the software ask learners to choose the meanings of target words?
(5) Productive recall: Does the software ask learners to produce the target word forms corresponding to the meanings provided?
(6) Productive recognition: Does the software ask learners to choose the target word forms corresponding to the meanings provided?
(7) Increasing retrieval effort: For a given item, does the software arrange exercises in the order of increasing difficulty?
(8) Generative use: Does the software encourage generative use of words, where learners encounter or use previously met words in novel contexts?
(9) Block size: Can the number of words studied in one learning session be controlled and altered?
(10) Adaptive sequencing: Does the software change the sequencing of items based on learners' previous performance on individual items?
(11) Expanded rehearsal: Does the software help implement expanded rehearsal, where the intervals between study trials are gradually increased as learning proceeds?

Downloaded by [Victoria University of Wellington] at 21:51 25 September 2012

Results

The nine flashcard programs were analysed using the 17 criteria described in the previous section. The results are summarised in Table 2.

Flashcard creation and editing

First, the nine programs were evaluated regarding their flashcard creation and editing functions. Table 2 shows that all the programs except *Word Engine* and *LearnThatWord* allow learners to create their own flashcards. The lack of a flashcard creation function is not necessarily a major shortcoming for these two programs because a large number of readymade flashcards are available for both programs. Users of *Word Engine* can purchase a wide selection of flashcard sets for English vocabulary learning (e.g. *Basic English, Advanced English*, or *TOEFL Vocabulary*) from the developer. *LearnThatWord* comes with an internal database of English vocabulary, which contains over 150,000 items (R. Warda, personal communication, 2 August 2010). However, learners who want to study technical vocabulary or a language other than English would require a flashcard creation function.

Multilingual support is beneficial not only because it allows learners to study various non-alphabet-based languages but also it contributes to improved performance using flashcards (Lado et al., 1967; Laufer & Shmueli, 1997; Mishima, 1967). All the programs with a flashcard creation function offer multilingual support. Using these programs, learners can create both target items and their translations in any language as long as a Unicode-compatible font is available.

Due to the increasing recognition that multi-word units play a major role in language acquisition and use for L2 learners (Pawley & Syder, 1983; Wray, 2000, 2002), an ideal program would let learners create flashcards for multi-word units as well as single words. As Table 2 indicates, all the seven programs with a flashcard creation function allow learners to do this.

Since vocabulary acquisition involves much more than associating new L2 words with their meaning (e.g. Beheydt, 1987; Nation, 1990, pp. 29–50, 2001, pp. 23–59), it is desirable that various types of information can be added to flashcards besides the word meanings. Once again, all the programs with a flashcard creation function allow learners to add various kinds of information about target words such as collocations, contexts, or images. In particular, *SuperMemo* and *vTrain* stand out among the others because they can handle more types of information than any other program.

Computers can help flashcard creation by automatically supplying information about lexical items. Unfortunately, only four of the nine programs offer support for flashcard creation. Among them, *iKnow!* provides the most comprehensive support. The program enables learners to import meaning, parts of speech, contexts and audio recordings of target words from existing flashcards created by publishers, language schools, or other users. *iKnow!* is also linked to Flickr.com (http://www.flickr.com/) and displays a list of images that are possibly related to the target word. Learners can choose an image representing the meaning of the word from the list and add it to their flashcard. It would also be valuable if software could automatically supply frequency information derived from a corpus or give a list of words in the same word family. Unfortunately, none of the programs is designed to automatically give frequency information or other words in the same word family. This will be a useful addition to future versions.

The last evaluation criterion regarding flashcard creation and editing is flashcard sets. Ideal flashcard software would allow learners to create their own sets of flashcards so that learners could review words belonging to the same semantic category (e.g. Beheydt, 1987; Nation, 2001, p. 103; Stahl & Nagy, 2006, pp. 77–96). All the programs except *Word Engine* have this function.

Learning

Next, the nine flashcard programs were analysed regarding their capabilities to facilitate learning. An ideal program would consist of two modes: a presentation mode, where learners familiarise themselves with the target words, and a retrieval mode, where they practise retrieval of previously met words. Table 2 shows that most programs are very similar as far as these two modes are concerned. All the programs except *Word Engine* and *LearnThatWord* have a presentation mode, and all nine programs have a retrieval mode.

Retrieval practice can be categorised into four types: receptive recall, receptive recognition, productive recall and productive recognition (Laufer et al., 2004; Laufer & Goldstein, 2004). Previous research suggests that flashcard software should support more than one type of retrieval practice. With the exception of *LearnThatWord*, all the programs provide at least two kinds of retrieval practice. *LearnThatWord* supports only the productive recall format probably because it assumes that learners already have receptive knowledge of target words, and emphasis is placed on acquiring the correct spelling.

Downloaded by [Victoria University of Wellington] at 21:51 25 September 2012

All the programs except *LearnThatWord* support multiple-choice exercises. However, *SuperMemo, vTrain* and *WordChamp* are limited in their capabilities in two respects. First, distractors are not automatically generated by these programs and need to be created by learners (in case of readymade flashcards, this is not necessary because distractors have been already created by flashcard authors). Second, these three programs present multiple choice options in a fixed order for a given item. This is not desirable because the position of the correct answer may offer inappropriate help in remembering. *MemoryLifter, P-Study System, Quizlet, iKnow!* and *Word Engine* not only generate distractors automatically but also change the position of the correct answer for each retrieval trial, providing a better condition for multiple-choice exercises.

The retrieval effort hypothesis implies that it is desirable for a flashcard program to arrange various types of exercises in the order of increasing difficulty (Bjork, 1994, 1999; Pyc & Rawson, 2009). Table 2 shows that only *WordChamp* and *iKnow!* are designed to gradually increase retrieval effort. *iKnow!* provides, in descending order of difficulty, the following five kinds of quizzes: receptive recognition without context, receptive recognition with context, productive recognition, productive recall with the spoken form provided and productive recall without the spoken form. The software also increases the number of multiple-choice options as learning proceeds. More specifically, when items are still unfamiliar to learners, *iKnow!* presents only five multiple-choice options. When learners become more familiar with the items, it gives 10 options, increasing the difficulty of retrieval practice. *Absolute Recall*™, a component of *WordChamp*, is also programmed to automatically increase retrieval effort. Yet, the program does not provide a wide variety of quizzes like *iKnow!*. More specifically, *Absolute Recall*™ offers only three kinds of quizzes, namely, receptive recall, productive recall and receptive recall with the spoken form provided.

In order to promote generative use (Joe, 1995, 1998; Nation, 2001, pp. 68–70), a good flashcard program would show the target word used in different senses, collocations, inflections, grammatical functions or sentence patterns every time the word is practised. Unfortunately, none of the programs supports generative use. It is true that using programs with a flashcard creation function, learners can create separate flashcards for different senses, collocations, or grammatical functions of a given item and study multiple aspects of word knowledge. However, this is not ideal for promoting generative use because if different aspects of a given item are treated as separate items, flashcard software has no control over which aspects of word knowledge are introduced in which order, and there is no guarantee that various aspects of word knowledge are systematically introduced to learners. All nine programs surveyed, therefore, fail to fully exploit computers' ability to present materials in an organised fashion to support generative use.

Studies have shown that the block size, or the number of items to be studied in one learning session, may influence flashcard learning (van Bussel, 1994; Kornell, 2009; Pyc & Rawson, 2007, 2009). Given the lack of consensus among researchers regarding the optimal block size, flashcard software should be flexible about the block size. As Table 2 shows, all the programs except *iKnow!* and *Word Engine* allow learners to study with a wide variety of block sizes.

Computers can facilitate systematic review of lexical items with an adaptive sequencing procedure (e.g. van Bussel, 1994; Nakata, 2008; Pyc & Rawson, 2007, 2009; Siegel & Misselt, 1984), or an algorithm to change sequencing of items based

Downloaded by [Victoria University of Wellington] at 21:51 25 September 2012

on learners' previous performance on individual items. Table 2 shows that all nine programs surveyed make use of adaptive sequencing procedures. Furthermore, all the programs except *Quizlet* and *LearnThatWord* support expanded rehearsal, where the intervals between study trials are gradually increased as learning proceeds. However, it should be noted that although many applied linguists as well as psychologists regard expanded rehearsal as the optimal learning schedule (e.g. Baddeley, 1997, pp. 112–114; Ellis, 1995; Hulstijn, 2001; Pimsleur, 1967), recent studies have revealed that it may exert a negative effect on learning in the long term (Cull, 2000; Karpicke & Roediger, 2007; Logan & Balota, 2008). Considering that expanded rehearsal is incorporated into many flashcard programs, further empirical studies on the review schedule are warranted.

Downloaded by [Victoria University of Wellington] at 21:51 25 September 2012

Discussion

Several observations can be made regarding the findings of the present analysis. First, the present investigation demonstrates that overall, there are discrepancies in the way flashcard programs are designed. The variations among the programs in their design suggest that there do not exist commonly accepted guidelines for how flashcard software should be designed. The evaluation criteria used in the present study may be a useful prototype of such guidelines.

Second, this study has suggested that in general, most programs have been developed in a way that maximises vocabulary learning. For instance, seven of the nine programs allow flashcard creation, offer multilingual support, support multi-word units, allow learners to add various kinds of information such as contexts, audios, or images to flashcards, have both retrieval and presentation modes, and are flexible about the block size. In addition, eight programs support flashcard sets and provide various types of exercises. Furthermore, all nine programs support scheduling.

Third, when individual programs are compared, *iKnow!* seems to be the best program among those investigated. It offers the most comprehensive support for data entry, automatically generates distractors for multiple-choice exercises, and increases retrieval effort by systematically introducing various types of exercises. Its only shortcomings may be that it limits the block size up to 10 and does not promote generative use of words.

Although *SuperMemo, VTrain, MemoryLifter* and *P-Study System* are very powerful, they are inferior to *iKnow!* in several respects. First, none of these programs arranges quizzes in the order of increasing difficulty, failing to gradually increase retrieval effort. Second, *vTrain* and *MemoryLifter* do not offer support for data entry. Third, *SuperMemo* and *VTrain* are not capable of automatically generating distractors for multiple-choice questions and do nothing more than to present learner-generated options in a fixed order. By addressing these drawbacks, these programs should contribute to more effective learning.

WordChamp, Quizlet and *Word Engine* are also limited in several respects. Flashcard creation is not supported by *Word Engine*, and expanded rehearsal is not supported by *Quizlet*. *WordChamp* does not automatically generate distractors for multiple-choice questions, and *Quizlet* and *Word Engine* are not designed to gradually increase retrieval effort. On a positive note, these three programs are equipped with an administrative tool, and teachers can keep track of students' progress including the number of flashcards created, studied or mastered. These

three programs may be ideal for teachers who would like to incorporate flashcard software into their courses.

LearnThatWord is a unique program because it assumes that learners already have receptive knowledge of target words, and emphasis is placed on acquiring the correct spelling. The software takes a different approach than others probably because it mainly targets English native speakers and ESL learners, who tend to have larger receptive vocabulary knowledge than EFL learners. According to the developer of *LearnThatWord*, new features such as multiple-choice quizzes or expanded rehearsal are scheduled to be introduced in the future (R. Warda, personal communication, 1 May 2010). However, they were not available when this study was conducted.

The present analysis has also shown that existing flashcard programs have some room for improvement. Most notably, as shown in Table 2, none of the programs is designed to encourage generative use of target words. It is unfortunate considering that generative use is regarded as essential for gaining deep understanding of target words (Joe, 1995, 1998; Nation, 2001, pp. 68–70) and that software can be designed to promote generation as one feature. Future software should support generative use by showing the target word used in different senses, collocations, inflections, or grammatical functions every time the word is practised.

Similarly, existing flashcard programs are limited in their ability to increase retrieval effort. The retrieval effort hypothesis (Bjork, 1994, 1999; Pyc & Rawson, 2009) implies that software should test learners' vocabulary knowledge in a relatively easy format in earlier stages and introduce a more demanding format later. Yet, only *WordChamp* and *iKnow!* arrange exercises in the order of increasing difficulty. *iKnow!*, which provides five kinds of quizzes and increases the number of multiple-choice options based on learners' memory states, is a good model of how to gradually increase retrieval effort.

Another possible area for improvement is support for data entry. The present study has shown that only four programs examined support flashcard creation by automatically supplying information about lexical items such as meaning, parts of speech, or contexts. Future programs should offer support for data entry because they can save learners the time-consuming task of flashcard creation and let them spend more time on studying. This can be achieved by allowing learners to import data from existing flashcards or a vocabulary database. This is an approach adopted by *WordChamp, Quizlet* and *iKnow!*. It should also be noted that although four of the nine programs offer support for flashcard creation, none of them automatically supplies frequency information derived from a corpus or gives a list of words in the same word family. Considering that frequency information gives learners a good indication of how useful a word is (Nation, 2001, pp. 6–22, 2008, pp. 7–15), the feature to provide frequency information will be a useful addition to future versions. Similarly, it would also be valuable if software could automatically give a list of words in the same word family because learning words from the same word family may help learners to efficiently increase their vocabulary size (Sökmen, 1992).

The last area for improvement is multiple-choice exercises. Although all the programs except *LearnThatWord* support multiple-choice quizzes, three of them do no more than to present learner-generated options in a fixed order. The ability to automatically generate distractors saves learners the task of creating distractors and

Downloaded by [Victoria University of Wellington] at 21:51 25 September 2012

should be incorporated into future software. Software should also change the position of the correct answer for each retrieval trial so that learners will not use it as an aid for remembering.

As pointed out earlier, computer-based flashcards offer benefits that paper-based ones do not (e.g. Ellis, 1995; Garcia & Arias, 2000; Hulstijn, 2001; Nation, 2001, pp. 108–110; Nesselhauf & Tschichold, 2002). The present investigation has suggested that overall, the nine programs surveyed exploit the advantages of computer-based instruction. Seven programs take advantage of the multimedia capabilities of CALL by allowing learners to add audios, images, or videos to flashcards. Moreover, all the programs make use of adaptive sequencing procedures, and seven of them support expanded rehearsal, which can be cumbersome if done manually. At the same time, the present study has also indicated that existing flashcard programs fail to exploit some advantages of computer-based instruction. For instance, only four programs offer support for data entry, two programs arrange exercises in the order of increasing difficulty, and none of them promotes generative use. Future software should exploit the potential of computer-based flashcards to the fullest in order to truly support vocabulary learning.

The present study has also shown the need for more studies on computer-based flashcards as well as flashcard learning in general. First, empirical studies have failed to identify the optimal block size or review schedule in flashcard learning (van Bussel, 1994; Cull, 2000; Karpicke & Roediger, 2007; Kornell, 2009; Logan & Balota, 2008; Pyc & Rawson, 2007, 2009). Further studies on these factors will have valuable implications for how computer-based flashcards should be designed. Second, the present study has only investigated software for PCs. An analysis of software developed for portable devices such as mobile phones, mp3 players, or game players (see Cobb & Horst, in press; Godwin-Jones, 2008, 2010, for examples of flashcard programs for portable devices) would be valuable. Lastly, it will be useful to investigate learners' reactions to computer-based flashcards to examine whether they accept flashcard programs developed according to learning principles.[1] Studies on learners' metacognition have suggested that learners tend to have misconceptions about what constitutes an effective learning technique (Bjork, 1994, 1999). For instance, many learners believe that massed learning is more effective than spaced learning although in fact, the opposite is true (Kornell, 2009; Kornell & Bjork, 2008). Furthermore, learners are often unaware that retrieval practice leads to superior retention than mere presentation (Karpicke, 2009; Karpicke, Butler, & Roediger, 2009). These findings imply that some learners may be unwilling to use research-based software due to lack of metacognitive knowledge. If learners feel uncomfortable with research-based programs, explaining the theoretical justification for software might be useful.

Downloaded by [Victoria University of Wellington] at 21:51 25 September 2012

Acknowledgements

My special acknowledgment goes to Stuart Webb, Kirsten Reid, Sky Marsen at Victoria University of Wellington, and anonymous reviewers for their invaluable advice. This study was supported by Student Exchange Support Program (Long-Term Study Abroad) Scholarship from Japan Student Services Organization.

Note

1. I thank an anonymous reviewer for pointing this out.

Downloaded by [Victoria University of Wellington] at 21:51 25 September 2012

Notes on contributor

Tatsuya Nakata is a PhD candidate at the School of Linguistics and Applied Language Studies, Victoria University of Wellington. His research interests are second language vocabulary acquisition and CALL. He is the winner of EUROCALL (European Association for Computer-Assisted Language Learning) Research Award for 2007.

References

Allum, P. (2004). Evaluation of CALL: Initial vocabulary learning. *ReCALL*, *16*(2), 488–501.

Baddeley, A.D. (1997). *Human memory: Theory and practice*. East Sussex: Psychology Press.

Bahrick, H.P., Bahrick, L.E., Bahrick, A.S., & Bahrick, P.E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, *4*(5), 316–321.

Bahrick, H.P., & Phelps, E. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *13*(2), 344–349. doi:10.1037/0278-7393.13.2.344.

Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning*, *52*(2), 323–363. doi:10.1111/0023-8333.00186.

Barcroft, J. (2004). Effects of sentence writing in second language lexical acquisition. *Second Language Research*, *20*(4), 303–334. doi:10.1191/0267658304sr233oa.

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.

Beheydt, L. (1987). The semantization of vocabulary in foreign language learning. *System*, *15*(1), 55–67. doi:0.1016/0346-251X(87)90048-0.

Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Bjork, R.A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 435–459). Cambridge, MA: MIT Press.

Böhlke, O. (2002). CALICO software review: VTrain 4.0 (Vocabulary Trainer). *CALICO Journal*, *21*(2), 458–469.

Burston, J. (2007). CALICO software review: WordChamp. *CALICO Journal*, *24*(2), 473–486.

Chapelle, C.A. (1998). Multimedia CALL: Lessons to be learned from research on instructed SLA. *Language Learning & Technology*, *2*(1), 21–39.

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

Chun, D., & Plass, J. (1997). Research on text comprehension in multimedia environments. *Language Learning & Technology*, *1*(1), 60–81.

Cobb, T., & Horst, M. (in press). Does Word Coach coach words? *CALICO Journal*.

Cull, W.L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, *14*(3), 215–235. doi:10.1002/(SICI)1099-0720(200005/06)14:3<215::AID-ACP640>3.0.CO;2-1.

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.

Ellis, N.C. (1995). The psychology of foreign language vocabulary acquisition: Implications for CALL. *Computer Assisted Language Learning*, *8*(2), 103–128. doi:10.1080/0958822940080202.

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

Foster, H. (2009). Building learner-generated vocabulary logs with Quizlet. *The Language Teacher*, *33*(12), 23–25.

Garcia, M.R., & Arias, F.V. (2000). A comparative study in motivation and learning through print-oriented and computer-oriented tests. *Computer Assisted Language Learning*, *13*(4/5), 457–465. doi:10.1076/0958-8221(200012)13:4-5;1-E;FT457.

Gillund, G., & Shiffrin, R.M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*(1), 1–67.
Godwin-Jones, R. (2008). Emerging technologies mobile-computing trends: Lighter, faster, smarter. *Language Learning & Technology*, *12*(3), 3–9.
Godwin-Jones, R. (2010). From memory palaces to spacing algorithms: Approaches to second-language vocabulary learning. *Language Learning & Technology*, *14*(2), 4–11.
Griffin, G.F., & Harley, T.A. (1996). List learning of second language vocabulary. *Applied Psycholinguistics*, *17*(4), 443–460. doi:10.1017/S0142716400008195.
Groot, P.J.M. (2000). Computer assisted second language vocabulary acquisition. *Language Learning & Technology*, *4*(1), 60–81.
Higgins, J. (1991). Which concordancer? A comparative review of MS-DOS software. *System*, *19*(1–2), 91–100. doi:10.1016/0346-251X(91)90010-M.
Hsieh, C., & Fei, F. (2009). Review of multimedia learning suite: Chinese characters. *Language Learning & Technology*, *13*(3), 16–25.
Hulstijn, J.H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal, and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge: Cambridge University Press.
Ishikawa, S. (2004). Vocabulary instruction at college level using JACET 8000 and educational software. In JACET Basic Words Revision Committee (Ed.), *How to make the best of JACET 8000: For educational and research application* (pp. 7–14). Tokyo: JACET.
Joe, A. (1995). Text-based tasks and incidental vocabulary learning. *Second Language Research*, *11*(2), 149–158. doi:10.1177/026765839501100206.
Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics*, *19*(3), 357–377. doi:10.1093/applin/19.3.357.
Karpicke, J.D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General*, *138*(4), 469–486. doi:10.1037/a0017341.
Karpicke, J.D., Butler, A.C., & Roediger, H.L. (2009). Metacognitive strategies in student learning: Do students practice retrieval when they study on their own? *Memory*, *17*(4), 471–479. doi:10.1080/09658210802647009.
Karpicke, J.D., & Roediger, H.L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 704–719. doi:10.1037/0278-7393.33.4.704.
Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology*, *23*(9), 1297–1317. doi:10.1002/acp.1537.
Kornell, N., & Bjork, R.A. (2008). Optimising self-regulated study: The benefits – and costs – of dropping flashcards. *Memory*, *16*(2), 125–136. doi:10.1080/09658210701763899.
Lado, R., Baldwin, B., & Lobo, F. (1967). *Massive vocabulary expansion in a foreign language beyond the basic course: The effects of stimuli, timing and order of presentation*. Washington, DC: US Department of Health, Education, and Welfare.
Laufer, B., Elder, C., Hill, K., & Congton, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing*, *21*(2), 202–226. doi:10.1191/0265532204lt277oa.
Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*(3), 399–436. doi:10.1111/j.0023-8333.2004.00260.x.
Laufer, B., & Shmueli, K. (1997). Memorizing new words: Does teaching have anything to do with it? *RELC Journal*, *28*(1), 89–108. doi:10.1177/003368829702800106.
Lexxica. (2010). Popular courses. *Word Engine*. Retrieved from http://www.wordengine.jp/courses
Logan, J.M., & Balota, D.A. (2008). Expanded vs. equal interval spaced retrieval practice: Exploring different schedules of spacing and retention interval in younger and older adults. *Aging, Neuropsychology, and Cognition*, *15*(3), 257–280. doi:10.1080/13825580701322171.
McNamara, D.S., & Healy, A.F. (1995). A generation advantage for multiplication skill training and nonword 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.

Downloaded by [Victoria University of Wellington] at 21:51 25 September 2012

Mishima, T. (1967). An experiment comparing five modalities of conveying meaning for the teaching of foreign language vocabulary. *Dissertation Abstracts International*, *27*, 3030A–3031A.
Modigliani, V. (1976). Effects on a later recall by delaying initial recall. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(5), 609–622. doi:10.1037/0278-7393.2.5.609.
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. doi:10.1017/S0958344008000219.
Nation, I.S.P. (1980). Strategies for receptive vocabulary learning. *RELC Guidelines*, *3*, 18–23.
Nation, I.S.P. (1982). Beginning to learn foreign vocabulary: A review of the research. *RELC Journal*, *13*(1), 14–36. doi:10.1177/003368282013001002.
Nation, I.S.P. (1990). *Teaching and learning vocabulary*. Boston: Heinle & Heinle.
Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
Nation, I.S.P. (2008). *Teaching vocabulary: Strategies and techniques*. Boston: Heinle Cengage Learning.
Nesi, H. (1996). Review article: For future reference? Current English learners' dictionaries in electronic form. *System*, *24*(4), 537–546. doi:10.1016/S0346-251X(97)88214-0.
Nesselhauf, N., & Tschichold, C. (2002). Collocations in CALL: An investigation of vocabulary-building software for EFL. *Computer Assisted Language Learning*, *15*(3), 251–279. doi:10.1076/call.15.3.251.8190.
Olmanson, J. (2007). Review of eSpindle Vocabulary & Spelling Program Online. *Language Learning & Technology*, *11*(3), 18–28.
Pawley, A., & Syder, F.H. (1983). Two puzzles for linguistic theory. In J.C. Richards & R.W. Schmidt (Eds.), *Language and communication* (pp. 191–226). London: Longman.
Pimsleur, P. (1967). A memory schedule. *Modern Language Journal*, *51*(2), 73–75.
Plass, J.L. (1998). Design and evaluation of the user interface of foreign language multimedia software: Cognitive approach. *Language Learning & Technology*, *2*(1), 35–45.
Pyc, M.A., & Rawson, K.A. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory & Cognition*, *35*(8), 1917–1927.
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.
Quizlet. (2010). About Quizlet. *Quizlet*. Retrieved from http://quizlet.com/about/
Rädle, P. (2009). VTrain (Vocabulary Trainer) – Awards. *VTrain.net*. Retrieved from http://www.vtrain.net/aw.htm
Rizo-Rodríguez, A. (2008). Review of five English learners' dictionaries on CD-ROM. *Language Learning & Technology*, *12*(1), 23–42.
Royer, J.M. (1973). Memory effects for test-like-events during acquisition of foreign language vocabulary. *Psychological Reports*, *32*, 195–198.
Salisbury, D.F., & Klein, J.D. (1988). A comparison of a microcomputer progressive state drill and flashcards for learning paired associates. *Journal of Computer-Based Instruction*, *15*(4), 136–143.
Siegel, M.A., & Misselt, A.L. (1984). Adaptive feedback and review paradigm for computer-based drills. *Journal of Educational Psychology*, *76*(2), 310–317.
Sökmen, A.J. (1992). Students as vocabulary generators. *TESOL Journal*, *1*(4), 16–18.
Stahl, S., & Nagy, W. (2006). *Teaching word meanings*. Mahwah, NJ: Lawrence Erlbaum.
Steinel, M.P., Hulstijn, J.H., & Steinel, W. (2007). Second language idiom learning in a paired-associate paradigm: Effects of direction of learning, direction of testing, idiom imageability, and idiom transparency. *Studies in Second Language Acquisition*, *29*(3), 449–484.
Tamaki, K. (2007, May 22). Incorporating Nintendo DS into the curriculum leads to marked improvement in English vocabulary: Yawata City Board of Education, Kyoto. *Mainichi Newspaper*, p. 28
Thorndike, E.L. (1908). Memory for paired associates. *Psychological Review*, *15*(2), 122–138. doi:10.1037/h0073570.

Downloaded by [Victoria University of Wellington] at 21:51 25 September 2012

Tinkham, T. (1993). The effect of semantic clustering on the learning of second language vocabulary. *System*, *21*(3), 371–380. doi:10.1191/026765897672376469.

Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research*, *13*(2), 138–163. doi:10.1191/026765897672376469.

van Bussel, F.J.J. (1994). Design rules for computer-aided learning of vocabulary items in a second language. *Computers in Human Behavior*, *10*(1), 63–76. doi:10.1016/0747-5632(94)90029-9.

Walker, B. (2006). VTrain 5.0. *TESL-EJ*, *10*(3). Retrieved from http://tesl-ej.org/ej39/m1.html

Waring, R. (1997a). The negative effects of learning words in semantic sets: A replication. *System*, *25*(2), 261–274. doi:10.1016/S0346-251X(97)00013-4.

Waring, R. (1997b). A study of receptive and productive learning from word cards. *Studies in Foreign Languages and Literature (Notre Dame Seishin University, Okayama)*, *21*(1), 94–114.

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

Webb, S.A. (2009a). The effects of pre-learning vocabulary on reading comprehension and writing. *Canadian Modern Language Review*, *65*(3), 441–470. doi:10.1353/cml.0.0046.

Webb, S.A. (2009b). The effects of receptive and productive learning of word pairs on vocabulary knowledge. *RELC Journal*, *40*(3), 360–376. doi:10.1177/0033688209343854.

Webber, N.E. (1978). Pictures and words as stimuli in learning foreign language responses. *The Journal of Psychology*, *98*(1), 57–63.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, *21*(4), 463–489. doi:10.1093/applin/21.4.463.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Downloaded by [Victoria University of Wellington] at 21:51 25 September 2012